Heat, Health and Central Valley Agricultural Populations: Examining Farmworker Vulnerability to Emerging Health Threats of Global Warming

Rachel Ng Jie Min

ABSTRACT

With increasing incidence of heat waves due to global warming, agricultural communities and farmworkers specifically will spend more work days under higher temperatures. The health impact of extensive time spent under heatwave conditions resulting in heat-induced illnesses are still unclear. To exacerbate these health related impacts, many farmworkers might not be able to reach infrastructure and public resources which can reduce vulnerability to heat illness. This project assesses the vulnerability of agricultural populations to the emerging health threats of global warming using a holistic framework by evaluating the incidence of heat induced illness in agricultural communities and the availability of healthcare resources. A linear regression weight analysis model was used for this analysis. The results show that while heat does not imply a direct correlation to possible heat induced illness, namely cardiorespiratory illness, it shows a possible stronger relationship to ethnicity and ages which relate to farmworker demographics. Healthcare resources were analyzed using ANOVA and T-test to understand the gap between agricultural counties and non agricultural counties while descriptive statistics were used to outline any variability. The results of this analysis highlight the insufficient healthcare resources available to agricultural communities. Another finding of this project is the need to standardize heatwave definitions to enable consistent and useful public policy for safeguarding the health of farmworkers. Unfortunately, this data modeling technique fails to encapsulate the socio-political demographic nuances of agricultural communities which is often abstracted from the data.

KEYWORDS

Central valley, agricultural communities, heatwaves, vulnerability analysis, healthcare resources, agricultural policy and regulation

INTRODUCTION

California is getting hotter - the 3 years before 2015 saw the hottest and driest years in California's instrumental record (Mann and Gleick 2015). In more recent times, 2020 was tied with 2016 to be the hottest year on record, ending the hottest decade ever observed. On a more global scale, global temperatures rose almost 2 ℉ from the average temperatures in the 20th century ("National Oceanic and Atmospheric Administration" n.d.). With these rising temperatures comes heat waves, characterized by more than 2 days with hotter than normal temperatures. In general, the last decade has witnessed more frequent occurrences of extreme heat and heat waves (Zuo et al. 2015). Over the past 2 years, the west coast has experienced heat waves consistently during the summer and early autumn months. In August 2020, Death Valley recorded a temperature of 130 ℉, cited as possibly the highest temperature ever reliably recorded on Earth ("California Heatwave Fits a Trend" 2020). During the same time, temperatures in Portland Oregon reached 116 ℉ in August 2021 (Press 2021). As the temperatures continue to rise due to global warming, present-day "heat wave" conditions may dominate summer months in the future (Miller et al. 2008). Heat waves will become more inbuilt into the narrative of west coast living, therefore the impacts of them need to be more strategically and rigorously analyzed.

Heat waves cause a variety of corresponding problems and impacts. Heat waves have significant impacts on issues ranging from health, infrastructure performance, energy demand, building design, water quality and overall cost of living. A case study from the 2009 heat wave in southern Australia showed that the impacts included 420 casualties, an estimated A$800 Million of financial losses derived from power outages, disruptions to public transport and general response expenditure (Zuo et al. 2015b). From a human health perspective, significant associations between ambient temperatures and respiratory disease mortality have been established (Zhao et al. 2019) and similarly, heatwaves have appeared to be marginally associated with cardiovascular morbidity (Cheng et al. 2019). Physiologically, in hotter temperatures heat must be transferred from the body to the environment and as such in order to maintain arterial blood pressure, cardiac output increases which in turn might increase people's vulnerability to cardiac illness. Clinical studies have also shown that individuals with preexisting heart disease are more susceptible to heat injury related to cardiac illness (Cui and Sinoway 2014). Similarly, clinical studies show that high temperatures incite vascular changes, release inflammatory mediators and decrease the effectiveness of immune responses; all of which directly influence respiratory morbidity. Furthermore, temperatures can also indirectly induce viral infections, bacterial activity or respiratory tract infections that result in respiratory illness (Zuo et al. 2015).

To exacerbate these health related impacts, many people might not be able to reach infrastructure and public resources which are known to reduce vulnerability to heat illness. For example, improved accessibility to air-conditioning could drastically reduce the chance of heat induced cardiorespiratory illness, but is not an accessible or viable option for everyone. The CDC advises the general public to maintain adequate hydration, seeking out air-conditioned buildings, wearing light-colored clothing and wide brimmed hats, and recommending the restriction of outdoor activities during heatwaves (CDC). However, some subsets of individuals, notable outdoor workers, by the nature of their work, may not be able to follow these guidelines to reduce their vulnerability to heatwave impacts. This paper focuses on agricultural workers - who by the nature of their outdoor work are faced with increasingly hazardous heat conditions due to extreme weather and hence heat waves. Farmworkers die of heat-related causes at roughly 20 times the rate of workers in any other civilian occupation (Ward 2010). Based on the current trajectory of rising temperatures, outdoor workers would work in temperatures about 100 ℉ 3 to 4 fold the number of days they currently do (Union of Concerned Scientists et al. 2021). Much of the available literature surrounding cardiorespiratory illness and heatwaves talks about mortality rate in general communities. However, urgent attention needs to be placed on subgroups who experience disproportionate growing risk as heat waves become more intense.

This project aims to assess the vulnerability of agricultural populations to the emerging health threats of global warming. First, I explored the variability of Californian weather to determine spatially how temperature, and specifically heatwaves, vary across the state. In particular, I evaluated if agricultural dense regions experience hotter days than no agricultural reasons. By focusing on periods of the year with hotter temperatures, I was able to focus on the effects of hotter periods on the cardiorespiratory emergency department visit rates. Finally, understanding the relationship between weather and ED visits is not complete without evaluating the availability of healthcare resources in different regions. Understanding the effects on temperature on cardiorespiratory ED visits alongside evaluating the available access to sufficient healthcare resources in different counties allows us to generate a full examination of general preparedness for emerging threats.

BACKGROUND

Defining Heat Waves And An Adaptive Model

Heatwaves are defined as more than 2 days with temperatures greater than 95 percentile for the community specific to 1st May to 30th September (Anderson and Bell 2011). By this definition, heat waves are characterized by the intense temperatures, duration and time during the season - specifically during the summer. However, this definition ignores the possibility of extreme heat outside the defined season and remains vague on the actual temperature that constitutes a heatwave as temperatures increase with climate change. Because of the diverse communities who are engaged in heat wave monitoring and research, there is no standard definition of a heat wave. Literature differs in threshold values, duration and ancillary values incorporated into heat wave definitions. While these diverse perspectives are important, the lack of a unified definition causes confusion when discussing patterns, trends, and impacts (Smith et al. 2013). Additionally, the lack of a specific definition of heatwaves is also caused by adaptive differences in population where, for example, 100 degree ℉ in Las Vegas might not be intolerable to the population there but would be considered a heat wave for populations in the cooler regions of Seattle. To account for the diversity in the definition of heat waves, I aim to develop an adaptive model for locating heat waves in this paper. To do this I redefined the scope for identification of heat waves as the following: The temperature of a region, defined on a zip code level, being greater than the 95th percentile of all maximum daily temperatures measured in the previous calendar year. This definition allows for regions that are historically cooler to have heat waves defined to be slightly cooler than historically hotter areas. Because this definition is adaptive, the algorithm would need to dynamically draw data to determine if heat wave conditions are met.

Central Valley

The Central Valley of California is a vast agricultural region which covers about 20,000 mi_2 of land. The valley is split into two large sections - the northern half being the Sacramento Valley and the southern half being the San Joaquin Valley (Faunt and Geological Survey (U.S.) 2009). The valley is an important agricultural region of the US, using fewer than 1% of US's farmland but providing 8% of the US agricultural output (USGS). Although the region is infamous for its profitable agricultural industry, it is also known for its endemic poverty and deep seeded social disparities, cultural diversity born of international migration and environmental pressures such as poor air quality (DeLugan et al. 2011). Demographically since 1990, there has been a growth in Hispanic, Black and Asian populations in the valley who now make up more than 50% of the populus (USDA). Economically, the central valley has a lower than state average per capita income, poverty rates and higher unemployment. In 2018, 13.8% of the jobs held in the central valley were related to farm work, 11.4% higher than the state average of 2.4% (CCSCE).

Summer days in the San Joaquin Valley are hotter than they were a century ago making heat waves more frequent (NOAA). This increase in frequency means that many of the agricultural workers have been waking up earlier to begin their harvest while still having to work in heatwave conditions and, often, haze from the California wildfire (Sainato 2021). Already, outdoor agricultural workers are feeling the intense heat and finding new ways to adapt to it. Fifteen years ago there was advocacy to create heat standards for California farmworkers after a spate of farmworker deaths, however nothing transpired and the United Farm Workers of America is still pushing for a national legislation to protect farmworkers from heat (Sengupta and Frank 2020). Therefore, this paper hopes to further draw attention to the impending temperature rises and consequently increased frequency of heatwaves - advocating for the urgency needed in locating and protecting farm workers who work tirelessly under the heat.

At these heatwave temperatures, the CDC recommends reduced working hours (CDC). However, for agricultural and construction workers, many of whom earn at or below minimum wage, reduced working hours also means reduced paychecks - which might not be a viable option (Sengupta and Frank 2020). The subgroup of outdoor workers also motivates the use of cardiorespiratory illness as the analysis metric for emergency department visits. Moving away from looking at heat exhaustion or heat stroke extends the concern away from merely drinking more water or taking a short break in the shade ("Effects of Heat - Climate and Human Health" n.d.). Instead, there is a physiological health impact that might not be immediately tangibly noticeable. Especially for outdoor workers, sudden cardiorespiratory illness that leads to emergency department visits would be a large cause of concern since it deviates from more traditional and known heat related health impacts, like heat stress and stroke, of working in hot conditions. Identifying areas with a high volume of outdoor workers is pivotal in ensuring that public policy ensures their safety during heatwave periods while also effectively allocating healthcare resources to regions with higher predicted amount of emergency department visits.

Why Do We Care About ED Visits?

Emergency Department (ED) visits can serve as an indication for cardiorespiratory vulnerability to heatwave conditions. Emergency department visits are mainly tracked by a collaborative network developed by CDC, state and local health departments, and academic and private sector health partners who collect electronic health data in real time in the National Syndromic Surveillance Program's (NSSP) database. The data included total ED visit volume, patient age, sex, region and most importntly reason for visit (Hartnett et al. 2020). Although literature has shown significant association between heatwaves and cardiovascular and respiratory mortality (Cheng et al. 2019), ED visits provide more insight into the scale at which heat waves affect people because they take into account non fatal incidents. ED visits are a good indicator of cardiorespiratory morbidity and discomfort caused by heat waves since they play an important role in wider access for seeking immediate care for serious conditions (Hartnett et al. 2020). Similarly, ED visits have been used to evaluate the access point for patients experiencing asthma exacerbations. ED visits allow for early identification, accurate assessment of the severity of airway obstruction before care decisions should be adjusted to the new severity level, leading to a disposition decision (Johnson et al. 2016). The use of ED in asthma escalation studies shows that they are a good indicator of initial symptoms and need for further care. Therefore, understanding how ED visits may increase in worsening heat wave conditions allows policy to focus on healthcare resource management to ensure sufficient safeguards are in place both to (1) reduce the risk of vulnerable populations from needing to visit the ED and (2) health resources are ready if ED visits increase.

Healthcare Resources

Like all resources, when healthcare resources are limited and demand exceeds supply, equitable allocation becomes an issue. When healthcare is constructed as a social good, allocation may proceed either in terms of competition between individuals on the basis of the relative strength of their competing rights, or on an aggregate basis by evaluating which distribution would produce the greatest amount of good for the greatest number of people (Kluge 2007). This creates an issue of balancing the competing rights and duties of healthcare providers with cost and outcome measures. In the context of healthcare resource allocation in and adjacent to agricultural communities, an inequitable distribution of resources might be more pronounced because of the relative purchasing power of the communities. Access to healthcare resources close to farmworkers is critical for maintaining their health. Traveling many miles, often an hour or more by public transit, is simply not an option for many farmworkers who work long hours and have limited access to transportation (Natsoulis et al. 2020). The three counties with the highest foreign-born uninsured populations, Fresno, Kern and San Joaquin, in the central valley have the worst access to healthcare for immigrants with the fewest clinics available per 1000 people (Natsoulis et al. 2020). These are also the most populated counties in the Central Valley.

METHODS

Site Selection Justification

This paper focuses on four different counties, Fresno, Tulare, San Luis Obispo and Santa Barbara - 2 within the Central Valley and 2 adjacent to the selected Central Valley sites (Figure 1). Fresno (Yellow) and Tulare (Green) were selected for being the top 2 agriculture counties in terms of net dollar earnings in the Central Valley (California Department of Food & Agriculture). San Luis Obispo (Blue) and Santa Barbara (Red) were selected as the counties outside of the central valley for their proximity to the selected counties. Additionally, to ensure maximum possible separation of selected sites for analysis, I ensured that the counties chosen from each group, within Central Valley and outside Central Valley, did not bother each other.

Figure 1: Selected sites on the map of California

Although this paper focuses on the Central Valley, selecting neighboring countries whose economic focus is not agriculture allows easier comparisons in future analysis, described below. Unlike Central Valley counties like Fresno and Tulare where a majority of the economy is fueled by agricultural outputs, a majority of San Luis Obispo's (SLO) employment is in government (San Luis Obispo, Chamber of Commerce), leisure and travel while Santa Barbara's is agricultural tourism and wineries - agriculture that focuses more on leisure like winery tours rather than industrial size crop production.

Dataset Development

I obtained my data from several repositories and this section includes a brief description of the dataset used. I included data collected from 2019 and not later to exclude any year including and after 2020 to avoid the possibility of COVID-19 skewing the data and results.

Global Historical Climate Network (GHCN) Datasets

The GHCN_daily dataset hosts data on daily land surface temperatures measured around the world by the National Center for Environmental Information under the National Oceanic and Atmospheric Administration (NOAA) (Datasets | Climate Data Online (CDO) | National Climatic Data Center (NCDC)). To obtain the relevant data, I made use of Google Big Query, where GHCN_daily is publicly available to extract data from California weather stations. The dataset also includes the names of all the GHCND stations available worldwide which allowed me to filter according to the stations in scope. The SQL query used is as follows:

This query filters the data down to the appropriate stations within the counties of interest and filters down to include only 2019 data.

California Department of Health Care Access and Information Datasets (HCAI)

Formally known as the Office of Statewide Health Planning and Development (OSHPD), HCAI has datasets about healthcare facilities, emergency department visits and reports on California Healthcare resources ("HCAI - Department of Health Care Access and Information"). The goal of HCAI is to make health information effective, affordable and accessible.

Current Healthcare facilities listing. The first dataset I included was HCAI's Current Healthcare facilities listing. This dataset includes all available and registered healthcare facilities in the state. This includes, but is not limited to, different facilities like General Acute Care Hospitals, Psychiatric Health Facilities, Skilled Nursing Facilities and Community Clinics.

Facilities Summary Reports. The facilities summary reports include summary statistics on Emergency Department visits including, the percentages of principal diagnosis and procedure groups per report period. The 4 different reporting periods were split as follows: January-March, April-June, July-September and October-December. Data on a finer grain where each row includes the exact date, demographic and diagnosis was not publicly available and hence this aggregated dataset had to be used instead. The relevant section was the summary of Principal Diagnosis, which included the following columns:

PRINCIPAL DIAGNOSIS GROUP
Certain conditions originating in the
Complications of pregnancy;
Congenital anomalies
Diseases of the blood and blood-
Diseases of the circulatory system
Diseases of the digestive system
Diseases of the genitourinary
Diseases of the musculoskeletal
Diseases of the nervous system
Diseases of the respiratory system
Diseases of the skin and
Endocrine; nutritional; and
Infectious and parasitic diseases
Injury and poisoning
Mental Illness
Neoplasms
No Default CCS
Residual codes; unclassified; all E
Symptoms; signs; and ill-defined

Table 2: Table including all principal diagnoses of Emergency Departments

For this research paper, I focused on Cardiorespiratory visits which would correspond directly to 'Diseases of the circulatory system' and 'Diseases of the respiratory system' in Table 2 above.

California Census Data - Quarterly Census of Employment and Wages (QCEW)

The QCEW program is a federal-state cooperative program which produces a comprehensive tabulation of employment and wage information for workers in California. The dataset includes the county, industry and number of employees which I used in this research paper to determine the total number of agriculture workers in the selected counties.

Analysis #1: Understanding The Californian Weather

The goal of this analysis was to designate heatwave days based on the definition of heatwaves described earlier - The temperature of a region, defined on a zip code level, being greater than the 95th percentile of all maximum daily temperatures measured in the previous calendar year. With this heat wave information, I can determine the period of times in the year where California is the hottest. This analysis would allow me to choose the ideal period that brackets these hotter days, January-March, April-June, July-September and October-December, that is the warmest to align with the time boundaries HCAI used to collect ED data. Additionally, this analysis would allow the tracking of differences between the temperatures across the 4 counties. For clarity and ease of identification, I labeled each period 1-4 as per the table below.

Period Label	1	2	3	4
Dates	01/01/19-03/31/19	04/01/19-06/30/19	07/01/19-09/30/19	10/01/19-12/31/19

Table 3: Period Label Correspondence

With these variations known, I used each period comparatively to look at statistical significance between the weather and ED visits.

The GHCN dataset did not include zip codes for the weather station used for measurement. Therefore the first step was to generate zip codes based on the longitude and latitude available in the GHCN dataset. To do this, I used GeoPy and the Nominatim API (Nominatim). Nominatim is an open source geocoding which converts lat/long into zip codes.

With the zip codes, I was then able to filter the data down to the counties of interest and group them by county. Because there were multiple weather stations in some counties, I took the daily average across all the weather stations to find the average minimum and maximum temperatures for each of the 4 counties. To understand the variability of the temperatures, I found the standard deviation across different factors: (1) SD of max temperature in each county across the whole year, (2) SD of max temperature across the 4 counties.

Using the average maximum temperature of each day and using the heat wave definition above, I found the number of days that would be classified as heat wave days in each county. I chose the 3 month period with the highest number of heatwave days as the hottest period as per the californian weather. As per my defined definition of heat waves, The temperature of a region, defined on a zip code level, being greater than the 95th percentile of all maximum daily temperatures measured in the previous calendar year, I had to measure the percentiles of temperature from 2018. To begin, I looked at the distribution of 2018 temperature grouped by county.

Analysis #2: Heat And Cardiorespiratory ED Visits

This analysis focuses on understanding the relationship between heat and cardiorespiratory ED visits. In this section, I performed all my analysis on aggregate data instead of grouping them into each county. This is because the dataset publicly available was too small. However, in looking at aggregate data across all the counties instead, I was able to analyze if variables related to Central Valley populations were more prominent in the prediction. To do this, I carried out 3 different statistical analyses.

Pearson Product-Moment Correlation

I plotted the number of incidences of individuals visiting the ED for Disease of the Circulatory System and Respiratory System against the period. Correlation is a measure of a monotonic association between 2 variables. A monotonic relationship between 2 variables is a one in which either (1) as the value of 1 variable increases, so does the value of the other variable; or (2) as the value of 1 variable increases, the other variable value decreases (Schober et al.). Because I am investigating the association between temperature and cardiovascular ED visits, the correlation coefficient (r) is a useful value. The r value is calculated with the following formula:

$$r = \frac{\Sigma(x - \mu_x)(y - \mu_y)}{\sqrt{\Sigma(x - \mu_x)^2 \Sigma(y - \mu_y)^2}}$$

$$x = One\ sample\ of\ temperature$$
$$\mu_x = Mean\ temperatures$$
$$y = One\ sample\ of\ \#\ ED\ visits$$
$$\mu_x = Mean\ \#\ ED\ visits$$

Linear Regression and Feature Importance

The goal of this section is to analyze if temperature is the primary driver for predicting the number of cardiovascular ED visits. To understand the relationship between various socio-environmental factors and ED visits, I built a LinearRegression model from the SKLearn module ("Sklearn.Linear_model.LinearRegression") in Python. However, the LinearRegression model may not have the best accuracy score and hence a RandomForest Regression model would also be trained to find which would offer a better accuracy score. To do this, a response variable and parameters will have to be chosen according to the following equation:

$$f_\theta(x) = \theta_0 + \theta_1 x_1 + \ldots + \theta_d x_d = \theta_0 + \sum_{j=1}^{d} \theta_j x_j$$

where **d** is the number of dimensions of the model, which follows the number of socio-environmental features the model has available for prediction. The response variable $f_\theta(x)$ is the variable the model aims to predict, in this case the number of ED visits. The additional socio-environmental factors added to the model were obtained from the US census data which includes variables like, the lat/long of the county, the average income, population size and other variables.

To build the model, I cleaned the dataset to ensure there is no missing or erroneous data that might skew the results. To better understand the model, Exploratory Data analysis and Principal Component Analysis (PCA) are carried out. Using these two techniques allowed me to identify outlier data and create a better domain understanding of the dataset. Furthermore, PCA allowed me to identify a possible low-rank data set, which would establish a clear link that a feature would capture most of the response variable variance.

Subsequently, I standardized and split the data into training and test sets. The training data set was used to build Linear Regression models. Different subset of features, based on features selected by the different techniques, will be used to train the models. To identify the model that is the most appropriate, I used iterations of cross validation with a 5 k-fold split to calculate the root mean square error (RMSE). RMSE would indicate the error in prediction, a small RMSE would indicate better predictions. Using the cross-validated dataset allowed me to maximize training accuracy without viewing the test set. The model with the lowest cross-validated RMSE was shipped and used for prediction and analysis.

An explanation for the model is carried out by looking at weights of the trained model to understand feature importances. The term feature importance describes how important the feature was for the classification performance of the model. This means that feature importance is a measure of individual contribution of a corresponding feature for the classifier (Saarela and Jauhiainen).

Analysis #3: Healthcare Facilities And The Outdoor Worker

This analysis explores the availability of resources across the 4 counties in relation to the number of outdoor workers in each of them. To do this, I carried out the two methods described below.

Spatial Autocorrelation and Moran's I

I used Moran's I (Sofianopoulou et al.) to investigate how closely values are clustered together in space, in this case in the 4 different counties selected. A high I value, where the possible values are between -1 and 1, indicates the variables of interest are more likely to be clustered together. A Moran's Index value near +1.0 indicates clustering, whereas an index value near −1.0 indicates dispersion. In my analysis, I analyzed 3 variables of interest, namely number of hospital beds, number of quarterly ED visits and number of healthcare resources.

To carry out calculation of Moran's I, I merged the compiled dataset which includes long/lat of different facilities and all the variables of interest with a shapefile of all the California counties. Using the libpysal and esda libraries available in python, I calculated Moran's I which has the following formula:

$$I = \frac{N\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i - \overline{x})(x_j - \overline{x})}{(\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij})\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

- Where:
  N is the number of observations (points or polygons)
  $\overline{x}$ is the mean of the variable
  X_i is the variable value at a particular location
  X_j is the variable value at another location
  W_{ij} is a weight indexing location of *i* relative to *j*

Descriptive statistics of facilities

This method generates the basic statistics for evaluation of the availability of healthcare facilities in the 4 different counties. The two subsets of descriptive statistics are sorting/grouping and illustrations and summary statistics (Kaliyadan and Kulkarni).

In the sorting/grouping subset, I looked at the total number of healthcare facilities grouped by different variables as exploration; these groupings included, by county and by facility type. In the summary statistics portion, I generated summary statistics which included the mean, median, max, minimum and standard deviation of all continuous variables available in the *Current Healthcare facilities listing* dataset. The first analysis conducted was looking at the number of beds available across the different counties normalized by the population size of the county. Number of hospital beds available is a common metric from previous reviews for understanding the optimal size and resource of hospitals (Giancotti et al.). Besides looking just at hospital beds available, I also accessed the number of different available facilities grouped by each county.

To establish if this difference was significant, I conducted an ANOVA test as well as pairwise t-test to find the p values. I used a threshold significance level of 0.1 instead of the usual 0.05 since increasing the threshold level lowers the evidentiary standard which is useful in the case of this paper since the dataset available was small. I conducted ANOVA to determine if the differences seen between the number of facilities was statistically significant. The basic idea behind a one-way ANOVA is to take independent random samples from each group, then compute the sample means for each group. After that, I compared the variation of sample means among the groups to the variation within the groups. Finally, make a decision based on a test statistic, whether the means of the groups are all equal or not. To ensure the data is well suited for ANOVA testing, I also conducted a normality assumption test using the residuals and a probability plot. I used the probability plot (Chambers et al., 1983) as a graphical technique for assessing whether or not my data set follows a normal distribution. The data are plotted against a theoretical distribution in such a way that the points should form approximately a straight line. Departures from this straight line indicate departures from the specified distribution.

RESULTS

Analysis #1: Understanding The Californian Weather

I first analyzed and explored the temperature data of all the measured data points from all weather stations in 2019. As shown in Figure R1.1, the maximum temperature measured follows a roughly normal distribution with a mean of 22.05 and a standard deviation of 8.63 (Figure 4). Because a lot of further analysis looked at information at a per county level, I additionally did analysis of temperature data on a by-county level, grouping all the information by the 4 different counties. Figure 5 and Table 6 below shows the individual distribution of temperature data per county along with a summary table of the mean and standard deviation of each distribution.

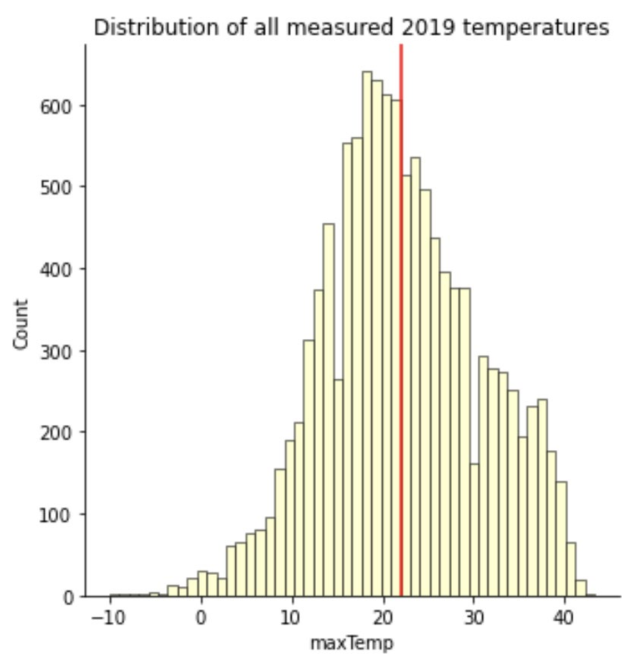

Figure 4: Distribution of temperature in 4 target counties

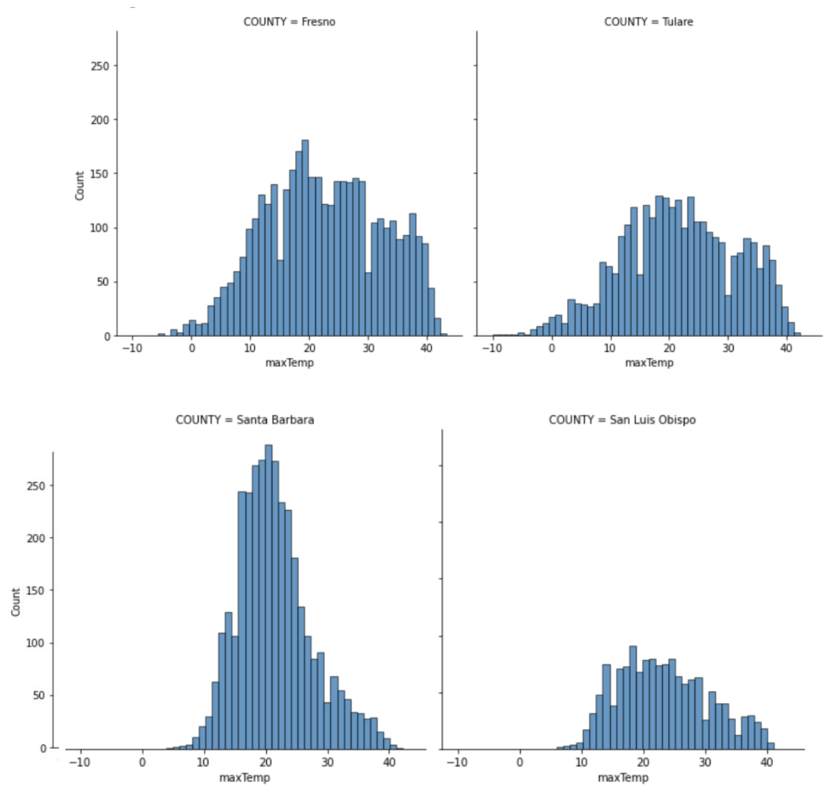

Figure 5: Distribution of temperature data

County	Statistic (2019)	
	Mean/°C	Standard Deviation/°C
Fresno	22.477	9.8733
San Luis Obispo	23.290	7.3693
Santa Barbara	21.431	6.0570
Tulare	21.56	9.7340

Table 6: Summary statistics for each county

Splitting the data per county reveals new temperature trends that were not accessible in the initial distribution (Figure 5). Interestingly, the 2 Central Valley counties, Fresno and Tulare experienced greater variability in temperature as shown in their larger standard deviations (Table 6). This higher standard deviation is attributed to the larger range of possible temperatures in the valley. Additionally, the distribution plots show that Fresno and Tulare experience a higher concentration of days with higher temperatures as seen in their distribution straying further away from normal compared to Santa Barbara. This means that the central valley has more days than have a higher maximum temperature compared to the non central valley counties.

Similar to the distribution above, the Figure 7 is grouped by county to see the variation across counties. In general, the graph shows that the hottest period of the year should fall around July to September. To get a more accurate reading on that, I analyzed the number of heat waves that happened in each period.

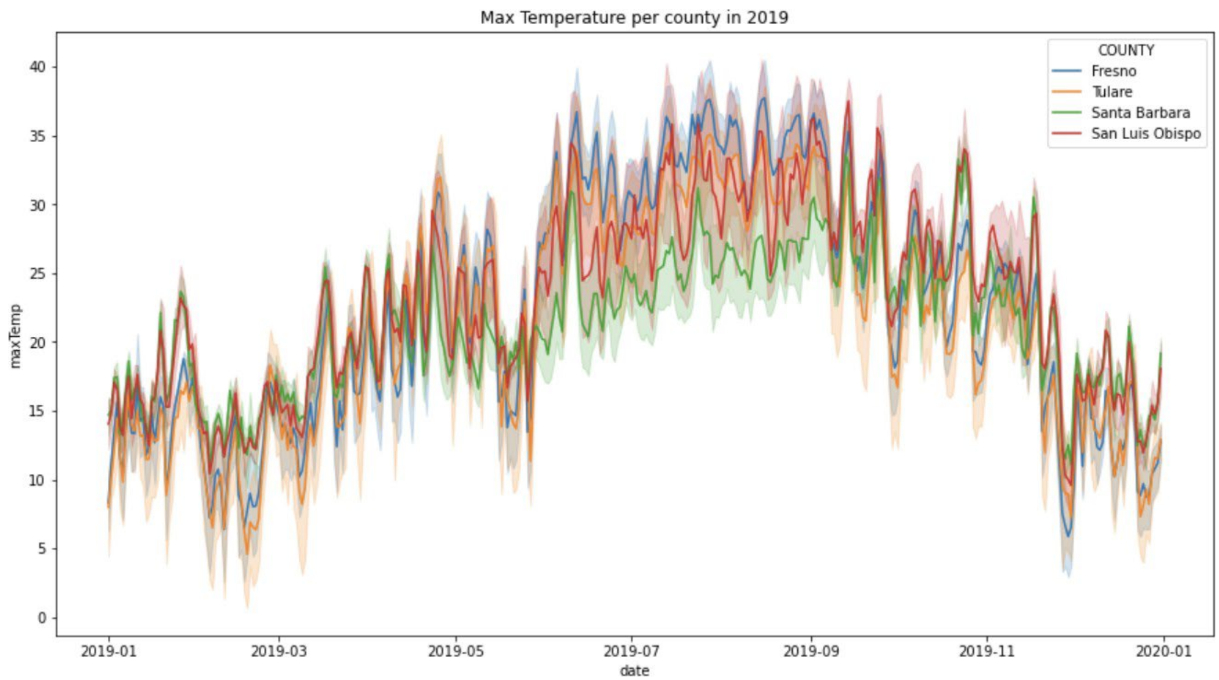

Figure 7: Time series graph of maximum temperature and variation in weather

To determine heat waves, I measured the percentiles of temperature from 2018 (Figure 8). To begin, I looked at the distribution of 2018 temperature grouped by county. Similar to 2019, the maximum temperatures for each county were roughly normal with San Luis Obispo and Tulare having a flatter distribution pulled to the right indicating more days with higher maximum temperatures. The calculated 95th percentile of maximum temperature for 2018 can be found in Table 9 below. The Central Valley states have a higher 95th percentile maximum temperature.

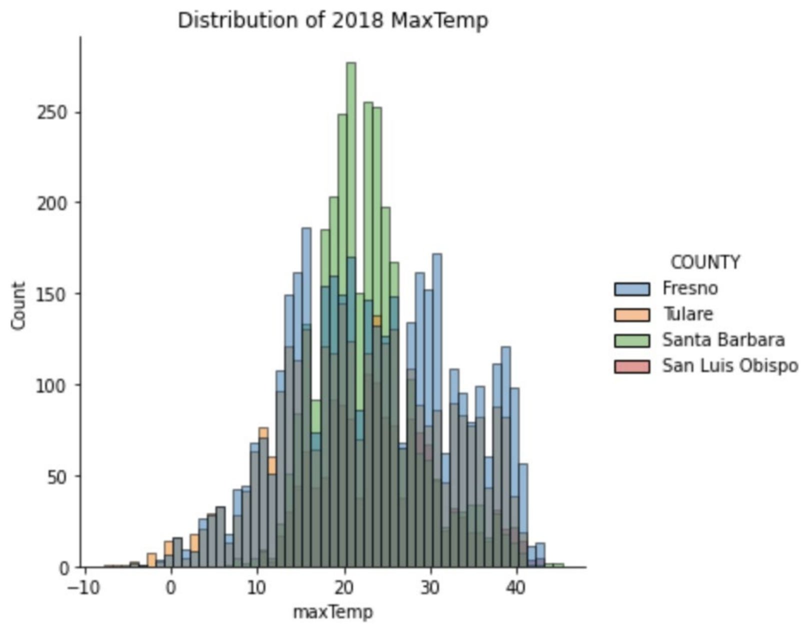

Figure 8: Distribution of max temperatures 2018

County	**95th Percentile maxTemp 2018/ °C**
Fresno	38.9
Tulare	37.8
Santa Barbara	34.4
San Luis Obispo	37.8

Table 9: Summary of 95th percentile of maximum temperature in 2018

		heatwave_days
COUNTY	**period**	
Fresno	**2**	14
	3	47
San Luis Obispo	**2**	5
	3	35
Santa Barbara	**2**	9
	3	58
	4	7
Tulare	**2**	11
	3	46

Table 10: Number of heatwave days per county per period in 2019

With the calculated 95th percentile, I counted the number of heatwave days in 2019 per county and per period. The results can be found in Table 10 above. With this table, I concluded that the hottest period for California weather is period 3, which aligns with 07/01/19-09/30/19 for having the most number of heatwave days that fell into this period across all 4 counties.

Analysis #2: Heat And Cardiorespiratory ED Visits

Pearson Product-Moment Correlation

The initial plots do not show a relationship between higher temperatures in period 2 and 3 and incidence of cardiorespiratory disease ER visits. In respiratory diseases however, the inverse relationship from the hypothesis was noted. Instead of higher ED visits during hotter periods, respiratory diseases saw higher incidence rates in period 1 and 4. This is likely because clinical discomfort due to respiratory illnesses can be exacerbated by cold temperatures like air conditioning or winter months which increases ED visits for respiratory illness (D'Amato et al.).

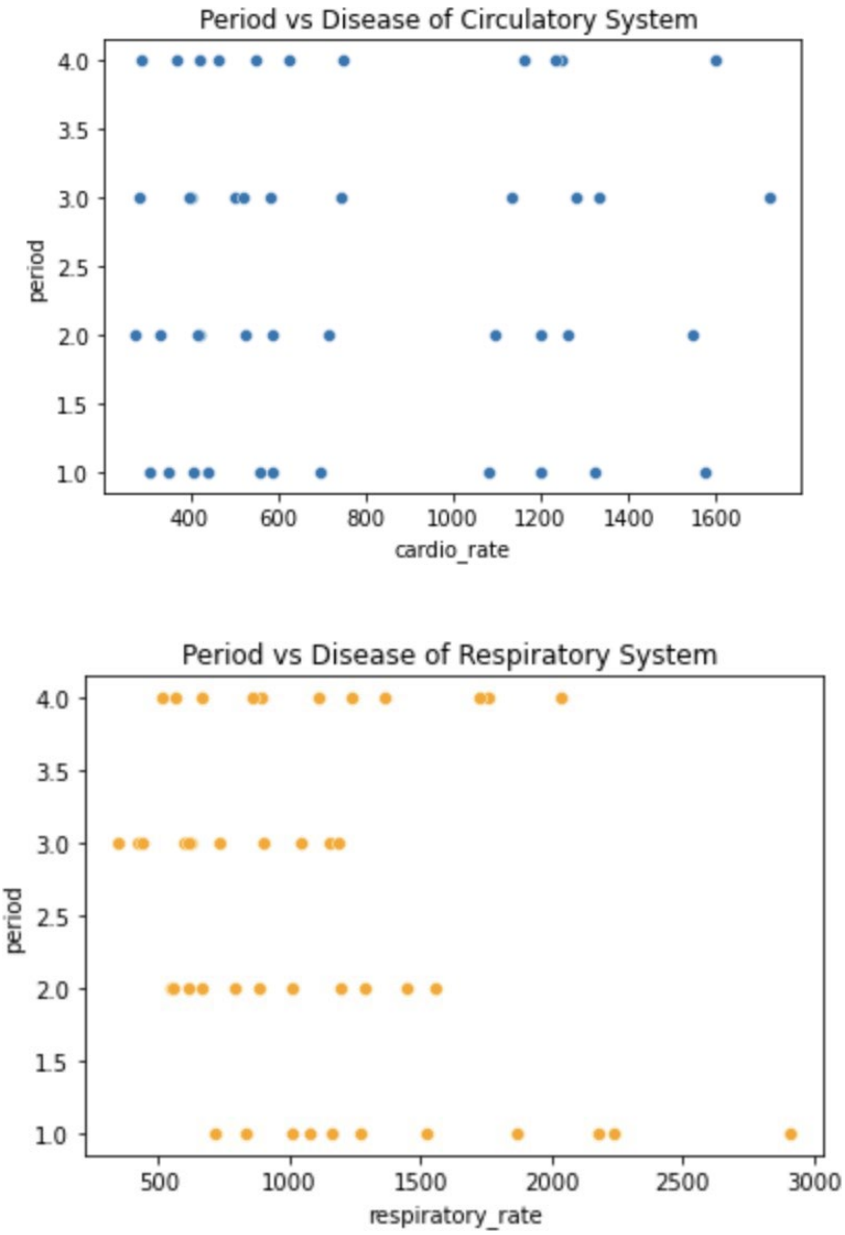

Figure 11: Incidence of disease of Cardiovascular system (Blue) and Respiratory system (orange) vs period

As seen in Figure 11, incidence of respiratory disease in the ER is most highly correlated to younger individuals aged, individuals of hispanic or latino descent and women. This aligns with the demographic of Central Valley populations who have a higher percentage of women and hispanic latino women as compared to non central valley populations (U.S. Census Bureau QuickFacts). This might be an indication that farmworkers experience higher incidences of respiratory illness in general. The variables in the Y axis of the plots are the possible parameters that relate to the rate of ED visits for each illness.

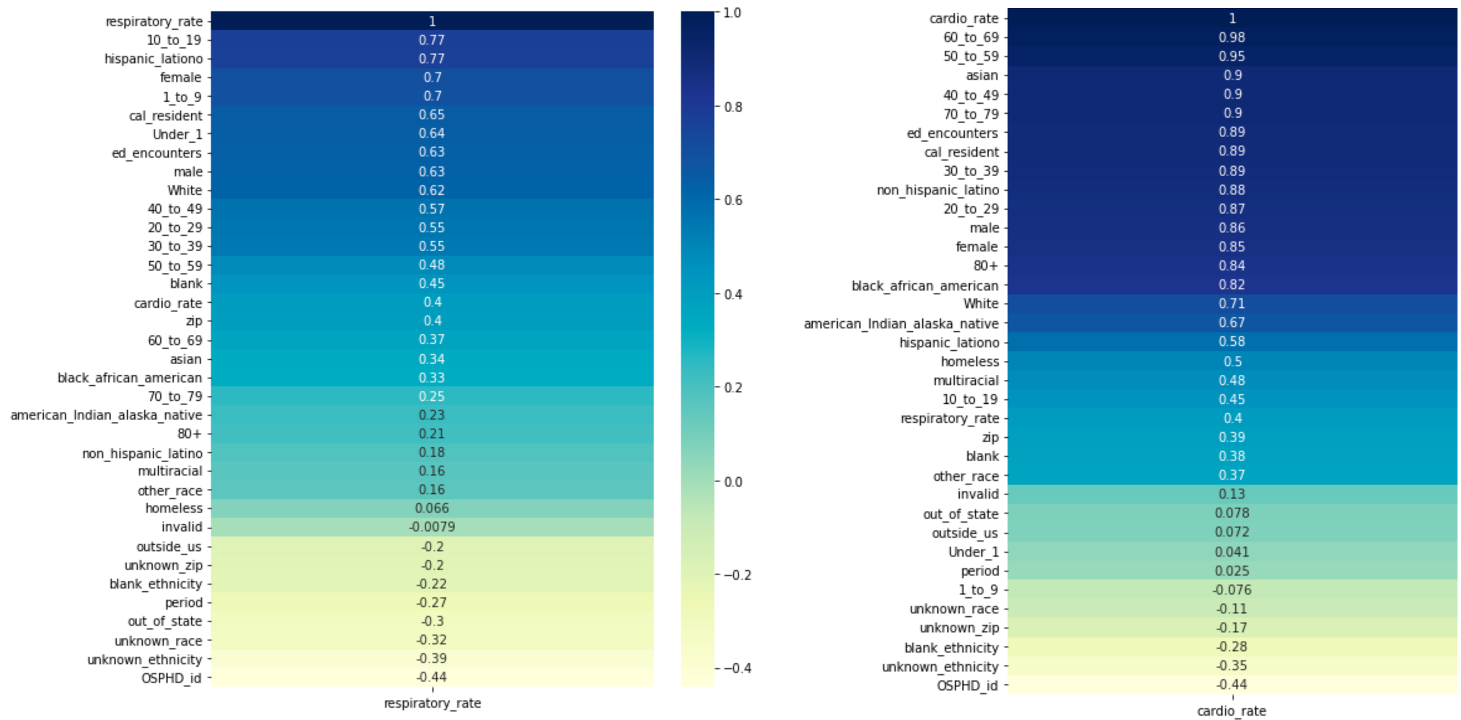

Figure 12: Correlation of variables to ED visits for respiratory and cardiovascular illness

Circulatory system illness, inversely, is more correlated to older individuals, asians and non hispanic latino individuals. However, correlation alone is not sufficient in providing evidence for understanding the relationship (Figure 12) . Therefore, I built a Linear Regression model to extract the weights of each variable in its predictive power of the response variable.

The first model is a linear regression model predicting the incidence rate of illness of the respiratory system. A graph of the weights trained for each feature can be found in Figure 13 below. In looking at the weights, I used the absolute value since both a negative and positive would contribute to the prediction of the response variable. In this case, the most important variables were leaving the ethnicity field blank, individuals from outside the US and individuals with unknown zip codes. In the instances where farmworkers visit the ED, these 3 factors align well with how they might respond to questionnaires. Therefore, this provides further indication that farmworkers might be more susceptible to respiratory illness in general regardless of the temperature. The trained model scored 0.9778 on the training data. This means that with the above weights in Figures R2.3, the model was able to correctly predict 97.78% of the sample training points again.

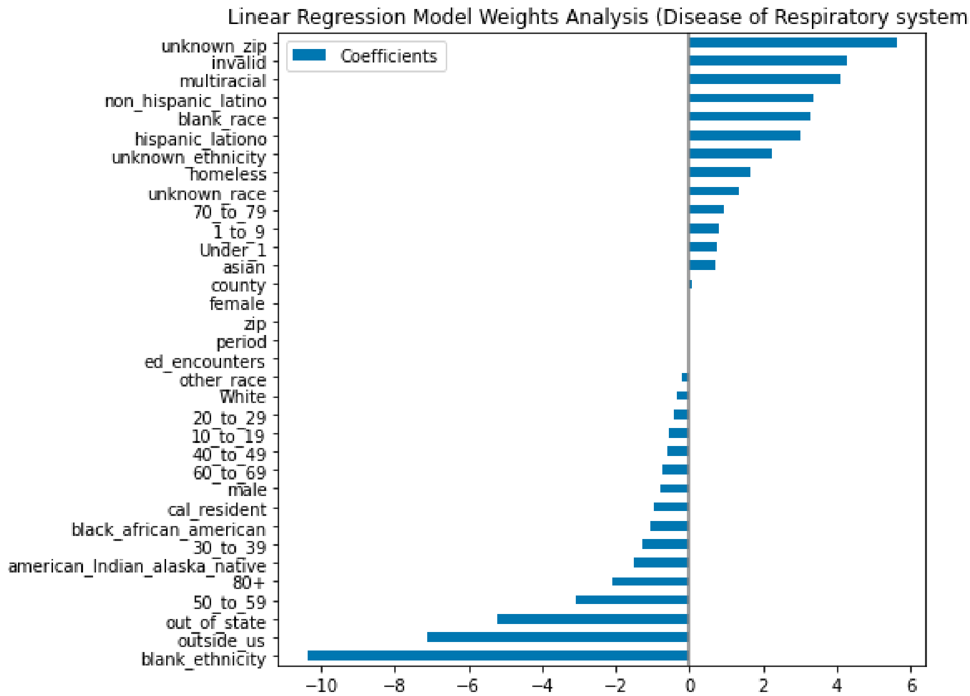

Figure 13: Weights trained in Linear regression model with disease of respiratory system as response

Figure 14 below shows the same weight bar graph but instead with disease of the circulatory system as the response variable. Similar to the correlation plot, older ages were a big contributor to the predictive power of the model. Additionally, a blank race field and American Indian alaskan natives field contributed the most to the model by having the largest weights with approximately 1 and 0.8 respectively. The above trained model scored 0.9964 on the training data. This means that with the above weights in Figures R2.3, the model was able to correctly predict 99.64% of the sample training points again.

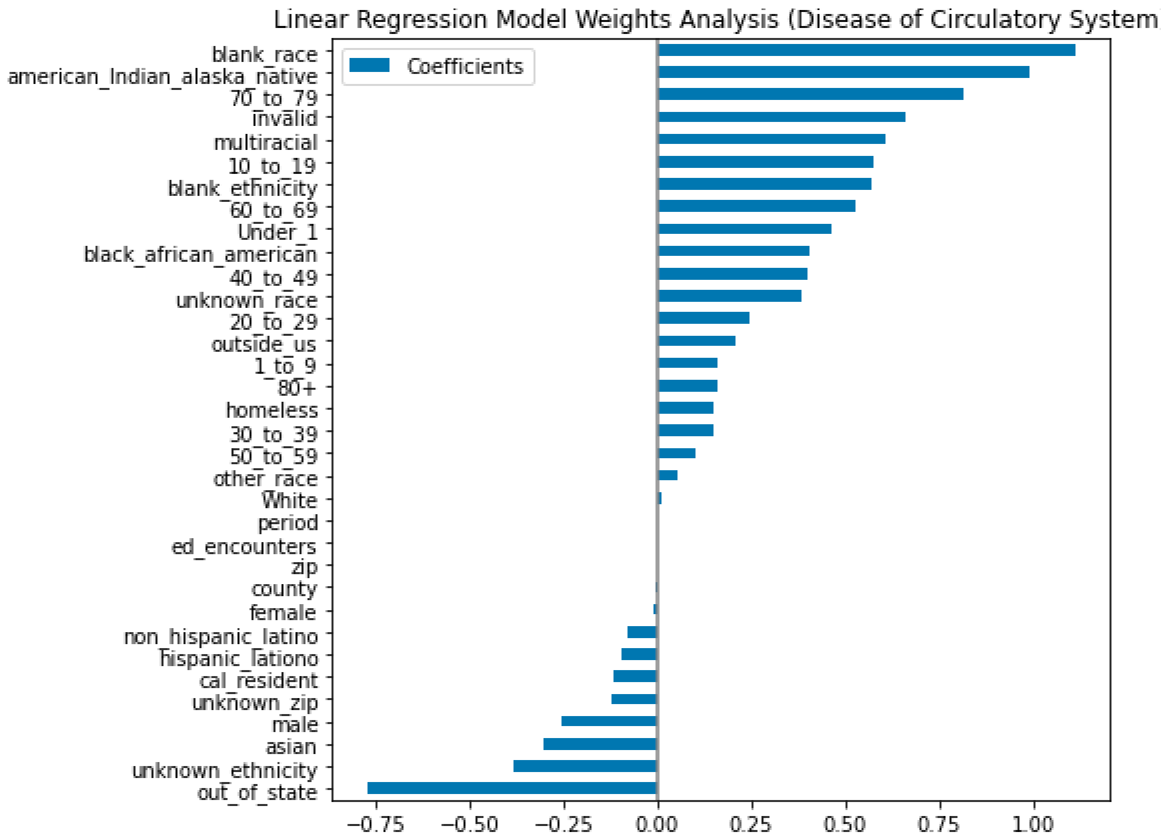

Figure 14: Weights trained in Linear regression model with disease of circulatory system as response

Analysis #3: Healthcare Facilities And The Farmworker

In terms of demographics from the census, Central Valley Counties have a larger Hispanic or Latino population, a higher percentage of individuals who do not speak English at home and a higher percentage of persons in poverty (Table 15). In addition, the 2 central valley counties have a much lower per capita income approximately $15,000 less compared to non central valley adjacent counties. This quick summary of census data across the counties aligns closely with the larger farmworker populations in Central Valley.

	Central Valley Counties		Non Central Valley Counties	
Fact	Tulare County	Fresno County	Santa Barbara County	San Luis Obispo County
Population Estimates, July 1 2021, (V2021)	477,054	1,013,581	446,475	283,159
Female persons, percent	50.00%	50.10%	50.00%	49.40%
Hispanic or Latino, percent	65.60%	53.80%	46.00%	22.90%
White alone, not Hispanic or Latino, percent	27.70%	28.60%	43.80%	68.50%
Language other than English spoken at home, percent of persons age 5 years+, 2016-2020	51.00%	44.20%	40.10%	17.10%
High school graduate or higher, percent of persons age 25 years+, 2016-2020	71.90%	77.30%	81.80%	91.80%
Bachelor's degree or higher, percent of persons age 25 years+, 2016-2020	14.50%	22.00%	35.00%	36.10%
Persons without health insurance, under age 65 years, percent	9.50%	9.70%	12.00%	7.00%
Median household income (in 2020 dollars), 2016-2020	$52,534	$57,109	$78,925	$77,948
Per capita income in past 12 months (in 2020 dollars), 2016-2020	$22,092	$25,757	$38,141	$38,686
Persons in poverty, percent	17.10%	17.10%	10.50%	10.60%

Table 15: Summarized census data of the 4 counties

From analysis 1, I established that the two Central Valley counties experience a high number of heatwave days and a flatter distribution of max temperatures indicating higher number of high temperature days. Furthermore, analysis 2 highlights the potential that Central Valley populations are more susceptible to diseases of the respiratory system. Given these compounding factors, it is important to ensure that Central Valley communities are well equipped with sufficient healthcare resources for the rate of illness and emergency department visits.

County	Total Number of Beds	Population Size	Beds/100,000 people
Fresno	6579	1013581	650
Tulare	2966	477054	622
San Luis Obispo	2572	283159	908
Santa Barabra	2248	446475	503

Table 16: Table showing total number of beds, population size, and beds/1000 people per county

All the counties reported higher beds/100,000 people than the North American average of 250 beds/100,000 people (Sen-Crowe et al.). However, Fresno, Tulare, and Santa Barbara counties had lower beds per 100,000 people compared to San Luis Obispo. Figure 17 shows the results of ANOVA's normality assumption check to ensure that the data provided is suitable for an ANOVA test.

The probability plot of the values indicated by the blue dots do not fall perfectly with the red line, this means there are slight departures from the normal distribution. However, in larger theoretical quantities the residuals after the model fit appear to have followed the normal distribution more closely. Given the smaller dataset, I assumed the normality to be true and carried on with the ANOVA test. The null hypothesis of the test is that the mean of the number of hospital beds across the counties should be equal.

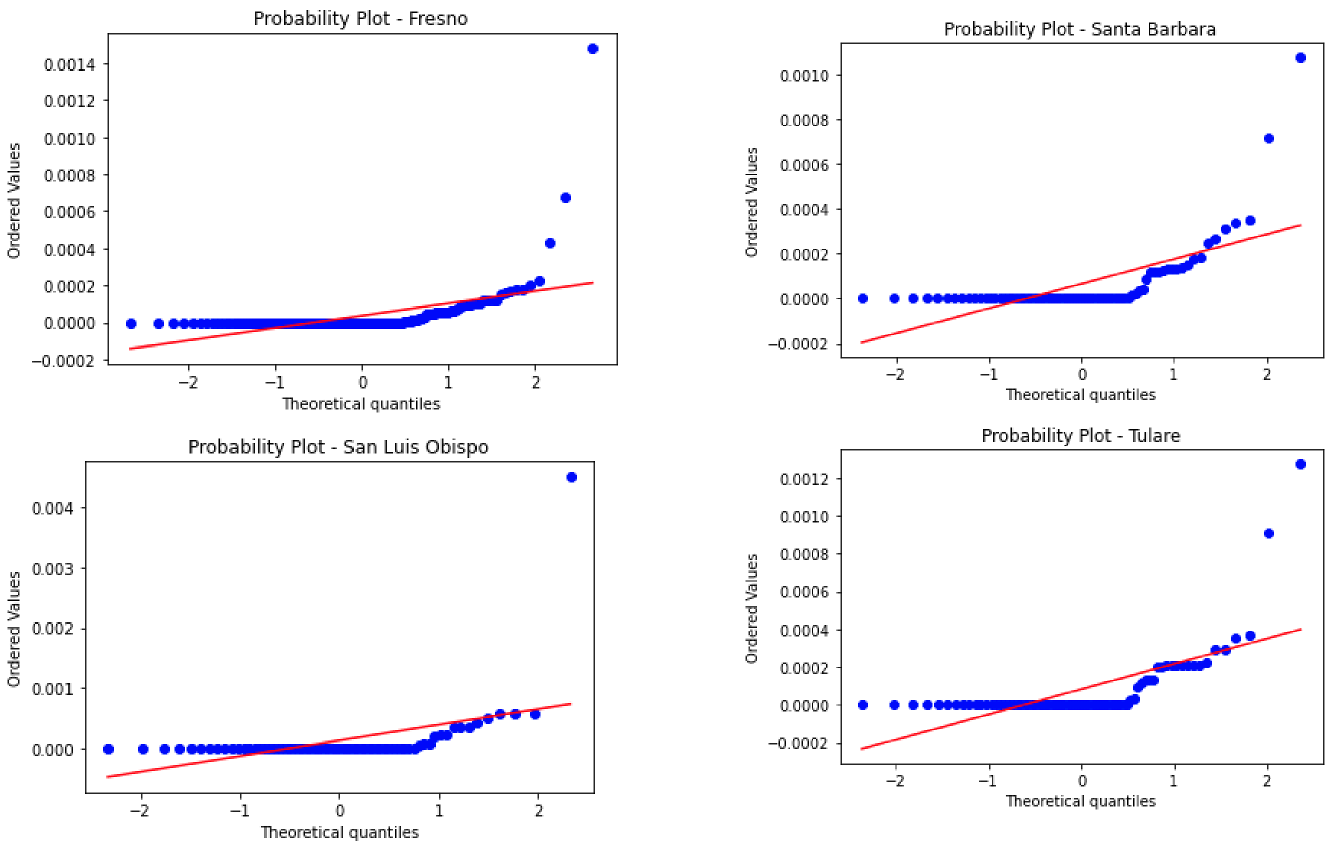

Figure 17: Probability Plots of Normalized number of beds per county

The result of the ANOVA test using stats.f_oneway is as follows:

ANOVA: F_onewayResult(statistic=2.204, pvalue=0.087).

Here, the p value is less than 0.1 and hence the difference in the null hypothesis is rejected. This implies that there is sufficient evidence to say that the difference between average number of hospital beds across counties is not a factor of chance at a 10% significance level.

A summary of the T and P values for each permutation can be found in Table 18 below.

COUNTY	Tulare	Fresno	Santa Barbara	San Luis Obispo
Tulare		T: 1.87 P: 0.0631	T: 0.7055 P: 0.4824	T: 0.5619 P: 0.575
Fresno			T: 1.415 P: 0.1612	T: 1.4042 P: 0.1628
Santa Barbara				T: 0.953 P: 0.3435

Table 18: Upper triangular matrix of T and P values from t-test

The p-value from the t-test between Tulare and Fresno was below the 0.1 threshold and hence it was statistically significant and hence we reject the null hypothesis that the difference in the means of hospital beds between the two countries was zero. This signifies that within Central Valley counties, there is significant difference in their access to healthcare resources measured in the number of beds available.

Similarly, a pairwise Tukey's Test shows that the difference in the mean between Fresno and San Luis Obispo was statistically significant. This is shown in Table 19 below.

```
        Multiple Comparison of Means - Tukey HSD, FWER=0.10
=====================================================================
     group1          group2     meandiff p-adj   lower  upper  reject
---------------------------------------------------------------------
         Fresno San Luis Obispo   0.0001 0.0598     0.0 0.0002   True
         Fresno   Santa Barbara      0.0 0.8335 -0.0001 0.0001  False
         Fresno          Tulare      0.0 0.5858    -0.0 0.0001  False
San Luis Obispo   Santa Barbara  -0.0001 0.4505 -0.0002    0.0  False
San Luis Obispo          Tulare     -0.0 0.6616 -0.0002 0.0001  False
  Santa Barbara          Tulare      0.0    0.9 -0.0001 0.0001  False
---------------------------------------------------------------------
```

Table 19: Tukey's Test

This conclusion is further supported by the calculated Moran's I value of 0.0183. This indicates that there is a small significance in the way facilities and their corresponding number of beds are clustered and that the spatial spread of the values are not perfectly random nor perfectly distributed. This means that facilities with a smaller number of beds could potentially be clustered in specific areas. Ideally, the Moran's I should be closer to -1 which would suggest perfect dispersal of facilities in space.

Figure 20 below shows the number of facilities available normalized by the size of the population.

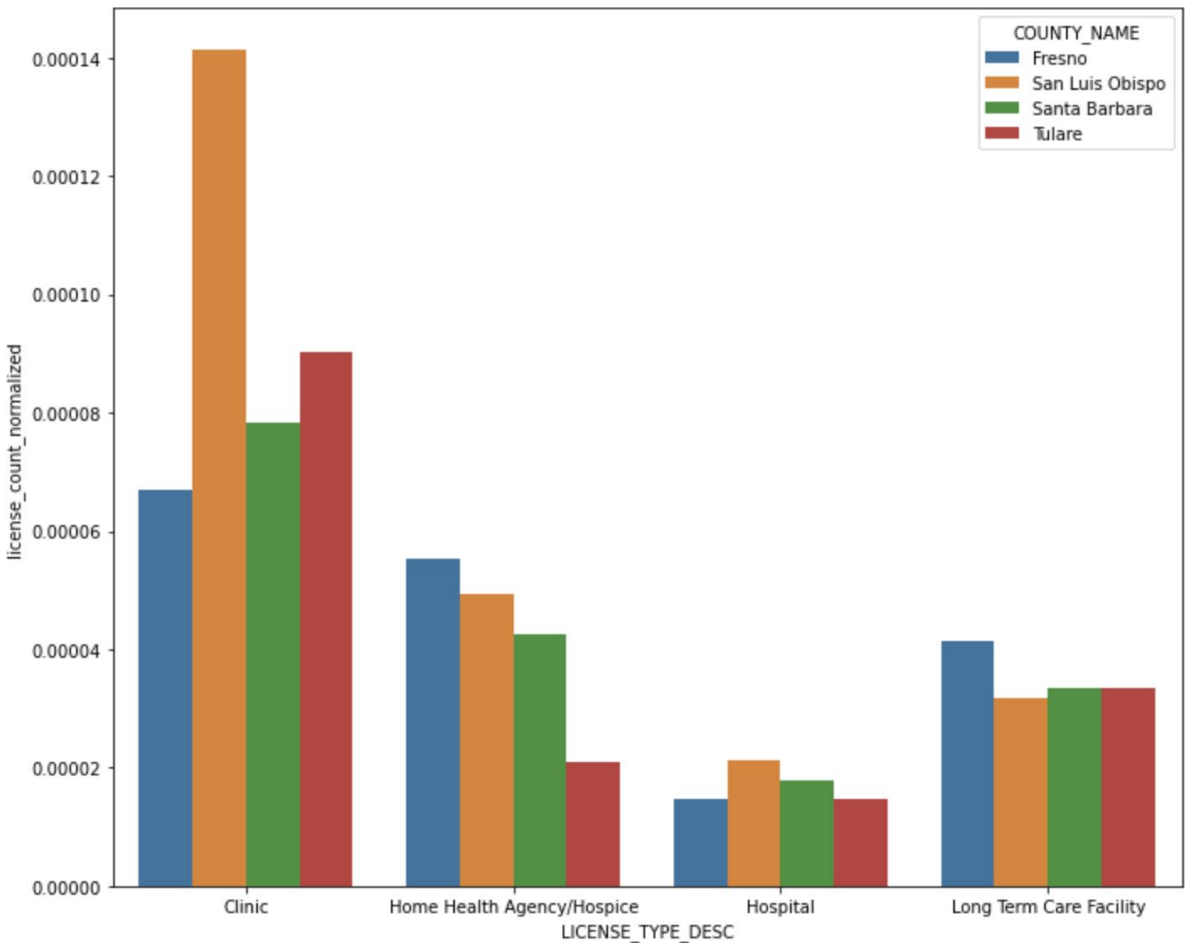

Figure 20: Breakdown of facilities available by county

Ideally, the normalized number of facilities available across the 4 counties should be approximately equal to indicate a proportionate distribution of healthcare resources. However, the bar graph shows the normalized number of hospitals available in Santa Barbara and San Luis Obispo were greater than in Fresno and Tulare (Figure 20). Hospitals provide greater access to comprehensive treatments compared to clinics and hence having a smaller number of them limits the access to healthcare. Additionally, San Luis Obispo is seen to have a disproportionately larger number of clinics as compared to the other 3 counties. Given the established vulnerability of Central Valley farmworkers who are at a 20 times rate of heat induced illness, the difference in availability of resources could be detrimental to the health of farmworkers.

DISCUSSION

With the food system being deeply connected to the climate and farm workers forming the basis of food production, the growing burden of increasing temperatures on food production begins to fall squarely on the farm worker. This research paper looked at heat and ED visit data as an indication of farmworker vulnerability as temperatures increase. However, the discussion of vulnerability is not complete without looking at other social and political factors that contribute to farmworker vulnerability in the central valley.

This section links the framing of farmworkers created in early agriculture policy to current psychosocial stressors that implicitly prevent farm workers from seeking healthcare which further put them at potential risk. These early policies which established the public image of farmworkers increased their social vulnerability, which antithetical diminishes the purpose of public policy as a necessary driver for protecting vulnerable communities. Through the lens of the historical and cultural context, this section criticizes the insufficiency of current regulation in protecting farmworkers from the emerging threat of heat-related illnesses with climate change. Updated regulations and policy needs to serve as the driving force in reducing the abstraction of the farmworkers from the food system, instead viewing farmworkers health as the basis of all agriculture policy.

Central Valley Farmworkers Demographics

Analysis #2 points to the potential of younger individuals of hispanic or latino descent being more susceptible to incidences of respiratory disease ER visits. This could point to a more intense relationship between race and respiratory disease ER visits because of the general demographic of Central Valley populations. A deeper analysis of the farmworker demographic reveals clear and intentional exploitation of their circumstances for cheap labor. 75% of farmworkers in the US are Latino migrants and about 50% of hired farmworkers do not have authorization to work in the US (Castillo et al. 2021), therefore in the eyes of government systems they are undocumented. In many states, undocumented individuals do not have access to government resouces - for example, they cannot obtain driver's licesnses, secure health coverge under the Affordable Care Act or apply for other public servives (Courville et al. 2016). This results in the data used in analysis #2 to be an underestimate of true Central Valley farmworker populations since many of these individuals are undocumentated and often abstracted from the data. This signals a more worrying trend that might be underestimating the vulnerability of farmworkers.

Historical Negligence Of Public Policy

Analysis #3 highlights the insufficient number of healthcare facilities to cater to the size of Central Valley populations. However, looking at historical public policy and the blatant negligence of farmworker health renders this result unsurprising. The lack of early policy that ensures sufficient resource availability of healthcare to farmworkers results in the difference in the number of hospital and general healthcare resources between Central Valley, who's major population compromises of farmworkers, and Non Central Valley communities seen in analysis #3. The following discusses how the set up of public policy has direct impact of the results we see in analysis #3.

Reviewing early public policy surrounding labor and environmental health clearly shows the prioritization of economic growth over labor welfare and health. The food system has always been interconnected to America's economy, growth and public policy. Despite the deep rooted interconnectivity and importance of the farmworker to the system, policy had still ignored worker health and instead incentivised economic accumulation and surplus in the growing agriculture industry. Policies that were enacted to ensure the economic and physical safety of workers, for example paying a livable wage and preventing over working, often excluded farmworkers as shown in the National Labor Relations Act and the Fair Labor Standards Act. The *National Labor Relations Act of 1935*, which guratenteeds the right of public sector employees to organize intro trade unions for collective bargaining and collective action had excluded agricultural workers (Cottle et al. 1983). The *Fair Labor Standards Act of 1938* which ensured workers be paid overtime if they worked more than a forty hour work week and a minimum wage also excluded farmworkers until the 1966 amendment that expanded coverage to some farmworkers. It was not until 1983 that an amendment was made to include migrant and seasonal agricultural workers. Policy intended to protect workers from harmful working environments had also conveniently excluded farmworkers. The *Clean Water Act (CWA)* is one such example. Established fully in 1972, CWA provided the basic structure for regulating discharge of pollutants into US water and regulated quality standards for surface water (US EPA 2013). However, the program exempts discharges associated with normal farming, ranching, and forestry activities such as plowing, cultivating, minor drainage, and harvesting for the production of food. (US EPA 2015). This allowed farming activities to pollute the environment which farmworkers worked in. In all three of the examples above, the policy prioritized productivity and output over their safety. The exclusion of farmworkers from these policies acknowledged their operational necessity in producing agricultural and hence economic output while explicitly ignoring their wellbeing. These policies were implemented with the knowledge of the demographic of farming communities, understanding their lower bargaining power and hence exploiting them for economic growth. The ruthlessness embedded in the policies aboves, where the health and livelihood of farm workers are willfully ignored, continues to be seen in today's policies.

Farmworkers, by the nature of the way big agriculture is set up, already face numerous environmental threats to their health. These threats include exposure to chemical hazards like pesticides, biological hazards like inadequate access to drinking water and finally physical hazards like the effects of heat exposure (Castillo et al. 2021). These environmental threats are exacerbated by psychosocial stressors like housing and food insecurity, discrimination and lack of social support. In addition, many farmworkers have to worry about their documentation status, the threat of deportation, language and cultural barriers. The conglomeration of these environmental, social and psychosocial stressors on farm workers synergistically make them a uniquely vulnerable population that has been exploited in public policy. Devastatingly, the same reasons they are exploited and excluded in policy is also a reason that their health might be at stake.

The treatment of migrant farmworkers during the Bracero Program Era highlights policymakers disregard for their health. The Bracero Program was a bi-lateral agreement between Mexico and the US that allowed millions of Mexican men to enter the US to work on short-term primarily agricultural labor contracts ("The Bracero Program " n.d.). Scholars have concluded that the Men who came to the US via the Bracero Program were exploited - their wages often unpaid and having little control of when and where to work. They relied on their employers for lodging and provision while the threat of deportation and the lack of effective mechanisms for challenging abuses left them vulnerable to contract violations and poor treatment (Toffoli 2018). From the viewpoint of policy, the Bracero Program was enacted to allow low wage labor to enter the US agriculture industry without fleshing out regulations that would protect the health and safety. Today, a similar structure exists in the US farmworker population where power dynamics between the employer and farmworkers reduces the bargaining power of the farmworker and puts them in vulnerable positions. The low bargaining power of farm workers is rooted in the many psychological stressors they constantly experience. Unsurprisingly, the implementation of aforementioned policies further increased the psychosocial stressors that farm workers experience by implicitly allowing farmworkers to be overworked and their health ignored. The continued ignorance of policy to the plight of farmworkers, seen through the lack of reporting channels to challenge abuse and systems to ensure their basic needs, reinforces capitalism's use and abuse of a dependent workforce underscored by a lacking public policy.

Insufficiency Of Current Policy For Rising Temperatures

The results of analysis #1 showcases warmer temperatures in Central Valley counties along with a flatter distribution where there are a greater number of hotter days in Central Valley counties compared to non Central Valley Counties. Furthermore, the analysis reveals a greater number of heatwave days in 2019 compared to 2018. This analysis, alongside the general trends of global warming, increases the need for targeted policy specifically for Central Valley agricultural workers' heat related safety. Despite this precedence, historical policy and updated regulation has been insufficient in protecting the health and safety of agricultural workers from hotter working conditions.

This historical set up of regulation and policy around the health and safety of farmworkers does not inspire confidence that current policy for heat related illness (HRI) will be sufficient. Currently there is no federal heat standard to lower the risk of HRI for farmworkers. The California Division of Occupational Safety and Health (Cal/OSHA) has been progressively adding regulations as more than 14 farmworkers died due to HRI between 2005 and 2015 (Langer et al. 2021). A summary of the regulations and recommendations can be found in table below:

Regulation/Recommendation	**Description**
Provision of water	Cool, no-cost and sufficient for each worker. Provide disposable cups. Encourage frequent drinking
Provision of shade	May be natural but not inside a building or car without AC. Enough for all employees.
Provision of rest periods	At least 5 minute break until all HRI symptoms are gone
High Heat provisions	Frequent communication to check alertness and symptoms
Acclimatization	New employees observed by supervisor for first 14 days
Training	Training covering signs, risks and symptoms of HRI. Training on emergency procedures.

Table 21: Table of Cal/OSHA HRI regulations (Langer et al. 2021)

A study of 587 farmworkers across 30 farms throughout the Central Valley found that HRI was exacerbated by work rate, measured in average heat rate of the farmworkers, environmental temperature and pay arrangement despite farms following Cal/OSHA regulations (Langer et al. 2021). The model of current Cal/OSHA may not be useful in farms where workers are paid by the piece, which is common for Central Valley's biggest commodities. Farmworkers may be hesitant to follow regulations as it may affect their bottom line as more time spent resting and drinking water means decreased earnings. Additionally, some farmworkers may be fearful that taking necessary breaks could lead to being questioned on their productivity and work ethic, which threatens their job security (Yeager n.d.). Therefore, future revisions need to first meet pace with increasing temperatures and also include practical recommendations targeted for specific tasks. Critically, regulations should be placed on supervisor behavior, who should model essential behavior since they are the main influence on farmworkers. Government processes and programs in this case should change power relations between those with privilege and those without, which encourages the exercise of political agency by the traditionally voiceless.

COVID-19 As A Indicator Of Preparedness For Emerging Threats

The analysis conducted throughout this paper has revealed the unpreparedness of healthcare resources and systems in protecting agricultural workers from heat related health issues. By viewing heat-induced illness as an emerging threat for agricultural worker's health, I likened the potential vulnerability of farmworkers to heat-induced illness to their vulnerability during the COVID-19 pandemic.

The recent and ongoing COVID-19 pandemic is a clear case study of accumulating psychosocial stressors impacting the health of farm workers. Undocumented individuals form a sizable population of essential farmworkers. However, their elevated risk of COVID-19 is compounded by their profession as farmworkers and their legal status as undocumented migrants. During the pandemic, essential workers were required by the federal government to continue work. Farmworkers, unsurprisingly, fell into this category of essential workers for their critical role in securing the nation's food supply and contributing to the global supply chain. Migrants and immigrants form 22% of all workers in the US food industry, therefore they play a pivotal role in the food system supply chain (Matthew et al. 2021). Yet, the same subset of individuals had the least protections like unemployment insurance, paid leave or healthcare coverage that would have been essential in safeguarding them as a workforce during the pandemic. Compounding on the insufficient safeguards for the farmworkers by the employer, farmworkers' low salaries also inhibit spending earnings on out-of-pocket healthcare costs while the need for daily earnings to sustain themselves and their families gives them little choice or incentive to get tested for fear of losing work. Additionally, the public health advice given during the pandemic, such as social distancing and quarantining were unattainable for farmworkers who had short windows for harvesting. The social vulnerability of farmworkers and the lack of a concerted effort to protect their health and safety amalgamates into multiple COVID-19 outbreaks at farms (Matthew et al. 2021). COVID-19 outbreaks at farms is not only a health issue for individual farmworkers but it also has disastrous impacts on the food system because of its disruption to the short and precise harvesting windows.

Similar to COVID-19, the health impacts of climate change can be seen as an emerging threat which farmworkers will be disproportionately affected by. Much of the available literature discusses the land suitability and crop yield implications of climate change. For example climate model predictions outline that changes in temperature and precipitation increases evapotranspiration and lower soil moisture levels while expanding the range and survival duration of agricultural pest populations (Schmidhuber and Tubiello 2007). Despite climate and environmental changes slowly reducing crop yield, consumers remain benefitting as real prices of food have been falling for the past 30 years ("Global food security under climate change" n.d.). The abstraction of the farmworker from climate change and food system discussions disregards the need to further regulation and policy which protect their health and safety, even as they continue to bear the burden of climate change impacts on food production. As climate change increases the number of days with high temperatures, the amount of time that farmworkers are exposed to extreme heat increases, in turn increasing their risk of heat-induced illnesses (Pan et al. 2021). Work schedules, shift rules, pay arrangements might increase farmworker exposure to extreme heat. For example, few alternatives might be available during periods of extreme heat while harvest have to be completed in the fields during specific timeframes that usually correspond to hot summer months. UCLA research also suggests that farmworkers working in extreme heat take home 5-10% less pay during 0.5-2 degree hotter-than-average years. This is a significant amount for potentially economically disadvantaged families (Yeager n.d.). Climate change stress on the larger food system eventually lands on the shoulders of farmworkers and as indicated by the response of COVID-19, the current safeguards for farmworkers are insufficient both for protecting individual health and the larger supply chain.

LIMITATIONS AND FUTURE DIRECTIONS

The biggest limitation of this research was the limited volume of data that I had been able to get access to. This makes sense because of the many protections around healthcare data that is necessary for privacy issues. Initially, I had reached out to the CDC in hopes of getting access to the Syndromic Surveillance dataset ("Overview | NSSP | CDC" 2021) termed the BioSense dataset. The dataset aims to have timely detection, understanding and monitoring of health events by tracking symptoms of patients in the emergency department. Access to the BioSense dataset would have given more granular information on ED resulting in more robust data analysis and identification of trends.

Using the HCAI data was only able to provide period-level aggregated information on ED visits. Additionally, a lot of the data inputted had been inconsistent across facilities and hence had to be removed, effectively reducing the size of the dataset. The prevalence of some heat induced illnesses are also not systematically recorded in the occupational setting and hence not quantified. This means that the results concluded in my analysis might be **underestimating the actual vulnerability of cardiorespiratory illness in the presence of heat**. Additionally, to account for the smaller dataset, I had also increased the threshold level for the p-value which lowers the evidentiary standard of the results. The smaller and aggregated characteristic of the dataset means that some of the analysis could not be as robust.

The final limitation was the heatwave definition that was used in the analysis to determine periods of highest heat waves. The definition had been a generalization of multiple other heatwave definitions from literature, therefore, if the analysis is repeated with a different definition of heatwave different results might ensue.

Further directions for this project would include obtaining more granular ED visit data. This would allow analysis into specific days which are classified as heatwave days instead of looking for periods of months that are determined to be the hottest since heatwave days can fall outside of the hottest period. With more granular data, future analysis would be able to enlist more sophisticated modeling techniques to determine which factors are the most important in predicting cardiorespiratory ED visits. Additionally, future analysis should incorporate standardizing the definition of heatwaves. With a standardized definition of heatwaves, policy and regulations for protecting farmworkers from heat will have a standard to follow. Currently, the multiple definitions of heatwaves could allow policymakers and regulatory agencies to skive extreme heat events as regular instead of classifying them appropriately as dangerous weather for farmworkers.

BROADER IMPLICATIONS

Under systems that operate under fundamental capitalist schemes, healthcare is largely a rhetoric of access and choice. However, the circumstances of many farmworkers as described earlier put them in socially vulnerable positions that revokes the perceived choice and reduces access to healthcare. Farmworkers are not only the most food insecure, experience biochemical threats and exposure due to the nature of farm work but also face unique psychosocial stressors that amplify their vulnerability. Historical public policy has not only been unsuccessful in protecting farmworker wellbeing but instead have amplified psychosocial stressors that may further exacerbate their risk of illness. As the basis of our food systems, farmworkers have ironically been abstracted away from discussion about the impact of climate change on the food production and supply chains. Climate change is an impending threat to the larger food system, the negative outcomes of which begins to fall squarely on the farmworkers. In general, stronger regulations are necessary to protect their health and wellbeing. Regulation updates need to meet pace with climate change while taking into account social vulnerabilities specific to farmworkers, such as legal status and power relations on the fields.

SPECIAL THANKS

Thank you to Patina and the 175B course staff for the constant support and reassurance throughout the process of writing this thesis. Thank you also to Professor Maureen for talking me through many of my ideas and confusions I had when planning out my thesis. Also for the many introductions and resources that you gave to get me going and the check ins to make sure I was on schedule and not too stressed. Thank you also to Emily Hernandez for working together and keeping me on track and accountable for my thesis. Thank you also to my family and friends - always supportive and always there!

BIBLIOGRAPHY

"HCAI - Department of Health Care Access and Information." *HCAI*, https://hcai.ca.gov/. Accessed 28 Apr. 2022.

"Sklearn.Linear_model.LinearRegression."*Scikit-Learn*, https://scikit-learn/stable/modules/generated/sklearn.linear_model.LinearRegression.html. Accessed 28 Apr. 2022.

Anderson, G. B., and M. L. Bell. 2011. Heat Waves in the United States: Mortality Risk during Heat Waves and Effect Modification by Heat Wave Characteristics in 43 U.S. Communities. Environmental Health Perspectives 119:210–218.

Anderson, G. B., M. L. Bell, and R. D. Peng. 2013. Methods to Calculate the Heat Index as an Exposure Metric in Environmental Health Research. Environmental Health Perspectives 121:1111–1119.

California Department of Food & Agriculture. *California Agricultural Statistics Review*. https://www.cdfa.ca.gov/Statistics/PDFs/2020_Ag_Stats_Review.pdf.

California Heatwave Fits a Trend. 2020, September 10. . Text.Article. https://earthobservatory.nasa.gov/images/147256/california-heatwave-fits-a-trend.

California's Central Valley | USGS California Water Science Center. (n.d.). . https://ca.water.usgs.gov/projects/central-valley/about-central-valley.html.

Castillo, F., A. M. Mora, G. L. Kayser, J. Vanos, C. Hyland, A. R. Yang, and B. Eskenazi. 2021. Annual Review of Public Health. Annual review of public health 42:257–276.

CCSCE - Center For Continuing Study of the California Economy. (n.d.). . http://www.ccsce.com/.
Chambers, John M., editor. *Graphical Methods for Data Analysis*. Wadsworth International Group; Duxbury Press, 1983.

Cheng, J., Z. Xu, H. Bambrick, V. Prescott, N. Wang, Y. Zhang, H. Su, S. Tong, and W. Hu. 2019. Cardiorespiratory effects of heatwaves: A systematic review and meta-analysis of global epidemiological evidence. Environmental Research 177:108610.

Cheng, J., Z. Xu, H. Bambrick, V. Prescott, N. Wang, Y. Zhang, H. Su, S. Tong, and W. Hu. 2019. Cardiorespiratory effects of heatwaves: A systematic review and meta-analysis of global epidemiological evidence. Environmental Research 177:108610.

Climate at a Glance | National Centers for Environmental Information (NCEI). (n.d.). . https://www.ncdc.noaa.gov/cag/statewide/time-series/4/tavg/ann/7/1895-2020?base_prd=true&begbaseyear=1901&endbaseyear=2000.

Cottle, R. L., H. H. Macaulay, and B. Yandle. 1983. Some economic effects of the California agricultural labor relations act. Journal of Labor Research 4:315–324.

Courville, M. D., G. Wadsworth, and M. Schenker. 2016. "We Just Have To Continue Working": Farmworker Self-care and Heat-related Illness. Journal of Agriculture, Food Systems, and Community Development 6:143–164.

Cui, J., and L. I. Sinoway. 2014. Cardiovascular responses to heat stress in chronic heart failure. Current heart failure reports 11:139–145.

D'Amato, Maria, et al. "The Impact of Cold on the Respiratory Tract and Its Consequences to Respiratory Health." *Clinical and Translational Allergy*, vol. 8, no. 1, May 2018, p. 20. *BioMed Central*, https://doi.org/10.1186/s13601-018-0208-9.

Datasets | Climate Data Online (CDO) | National Climatic Data Center (NCDC). https://www.ncdc.noaa.gov/cdo-web/datasets. Accessed 28 Apr. 2022.

DeLugan, R. M., M. D. Hernandez, D. E. Sylvester, and S. E. Weffer. 2011. The Dynamics of Social Indicator Research for California's Central Valley in Transition. Social Indicators Research 100:259–271.

Effects of Heat - Climate and Human Health. (n.d.). . https://www.niehs.nih.gov/research/programs/climatechange/health_impacts/heat/index.cfm.

Faunt, C. C. and Geological Survey (U.S.), editors. 2009. Groundwater availability of the Central Valley Aquifer, California. U.S. Geological Survey, Reston, Va.

General Technical Reports | Rocky Mountain Research Station. (n.d.). . https://www.fs.usda.gov/rmrs/publications/series/general-technical-reports.

Giancotti, Monica, et al. "Efficiency and Optimal Size of Hospitals: Results of a Systematic Search." *PLoS ONE*, vol. 12, no. 3, Mar. 2017, p. e0174533. *PubMed Central*, https://doi.org/10.1371/journal.pone.0174533.

Global food security under climate change. (n.d.). https://www.pnas.org/doi/abs/10.1073/pnas.0701976104.

Hartnett, K. P., A. Kite-Powell, J. DeVies, M. A. Coletta, T. K. Boehmer, J. Adjemian, and A. V. Gundlapalli. 2020. Impact of the COVID-19 Pandemic on Emergency Department Visits — United States, January 1, 2019–May 30, 2020. Morbidity and Mortality Weekly Report 69:699–704.

Heat Stress Recommendations | NIOSH | CDC. 2020, November 16. . https://www.cdc.gov/niosh/topics/heatstress/recommendations.html.

Johnson, L. H., P. Chambers, and J. W. Dexheimer. 2016. Asthma-related emergency department use: current perspectives. Open Access Emergency Medicine : OAEM 8:47–55.

Kaliyadan, Feroze, and Vinay Kulkarni. "Types of Variables, Descriptive Statistics, and Sample Size." *Indian Dermatology Online Journal*, vol. 10, no. 1, 2019, pp. 82–86. *PubMed Central*, https://doi.org/10.4103/idoj.IDOJ_468_18.

Karevan, Z., and J. A. K. Suykens. 2020. Transductive LSTM for time-series prediction: An application to weather forecasting. Neural Networks 125:1–9.

Kluge, E.-H. W. 2007. Resource Allocation in Healthcare: Implications of Models of Medicine as a Profession. Medscape General Medicine 9:57.

Langer, C. E., D. C. Mitchell, T. L. Armitage, S. C. Moyce, D. J. Tancredi, J. Castro, A. J. Vega-Arroyo, D. H. Bennett, and M. B. Schenker. 2021. Are Cal/OSHA Regulations Protecting Farmworkers in California From Heat-Related Illness? Journal of Occupational & Environmental Medicine 63:532–539.

Mann, M. E., and P. H. Gleick. 2015. Climate change and California drought in the 21st century. Proceedings of the National Academy of Sciences 112:3858–3859.

Matthew, O. O., P. F. Monaghan, and J. S. Luque. 2021. The Novel Coronavirus and Undocumented Farmworkers in the United States. NEW SOLUTIONS: A Journal of Environmental and Occupational Health Policy 31:9–15.

Miller, N. L., K. Hayhoe, J. Jin, and M. Auffhammer. 2008. Climate, Extreme Heat, and Electricity Demand in California. Journal of Applied Meteorology and Climatology 47:1834–1844.

National Oceanic and Atmospheric Administration. (n.d.). . http://www.noaa.gov/.

Natsoulis, Alizée, & Slootjes, Jasmijn. 2020. "Gaps in Health Services for Immigrants in the Central Valley". Berkeley, CA: Berkeley Interdisciplinary Migration Initiative.

Nominatim. https://nominatim.org/. Accessed 28 Apr. 2022.

Overview | NSSP | CDC. 2021, November 5. . https://www.cdc.gov/nssp/overview.html.

Pan, Q., D. A. Sumner, D. C. Mitchell, and M. Schenker. 2021. Compensation incentives and heat exposure affect farm worker effort. PLOS ONE 16:e0259459.

Press, T. A. 2021, August 11. Residents In The Pacific Northwest Are Getting Ready For Another Heat Wave. NPR.

Saarela, Mirka, and Susanne Jauhiainen. "Comparison of Feature Importance Measures as Explanations for Classification Models." *SN Applied Sciences*, vol. 3, no. 2, Feb. 2021, p. 272. *Springer Link*, https://doi.org/10.1007/s42452-021-04148-9.

Sainato, M. 2021, July 16. 'We're not animals, we're human beings': US farm workers labor in deadly heat with few protections. The Guardian.

San Luis Obispo, Chamber of Commerce. *San Luis Obispo Community and Economic Profile*. https://slochamber.org/wp-content/uploads/2020/02/2020-Community-Econ.-Profile-1.pdf.

Schober, Patrick, et al. "Correlation Coefficients: Appropriate Use and Interpretation." *Anesthesia & Analgesia*, vol. 126, no. 5, May 2018, pp. 1763–68. *DOI.org (Crossref)*, https://doi.org/10.1213/ANE.0000000000002864.

Sen-Crowe, Brendon, et al. "A Closer Look Into Global Hospital Beds Capacity and Resource Shortages During the COVID-19 Pandemic." *The Journal of Surgical Research*, vol. 260, Apr. 2021, pp. 56–63. *PubMed Central*, https://doi.org/10.1016/j.jss.2020.11.062.

Sengupta, S., and B. L. Frank. 2020, August 25. Heat, Smoke and Covid Are Battering the Workers Who Feed America. The New York Times.

Smith, T. T., B. F. Zaitchik, and J. M. Gohlke. 2013. Heat waves in the United States: definitions, patterns and trends. Climatic change 118:811–825.

Sofianopoulou, E., et al. "Use of Spatial Autocorrelation to Investigate Clustering of Health Deprivation:" *Epidemiology*, vol. 17, no. Suppl, Nov. 2006, p. S95. *DOI.org (Crossref)*, https://doi.org/10.1097/00001648-200611001-00228.

Summer 2020 ranked as one of the hottest on record for U.S. | National Oceanic and Atmospheric Administration. (n.d.). . https://www.noaa.gov/news/summer-2020-ranked-as-one-of-hottest-on-record-for-us.

The Bracero Program. (n.d.). https://www.labor.ucla.edu/what-we-do/research-tools/the-bracero-program/ .

Toffoli, E. 2018. Capturing Capitalism's Work. Radical History Review 2018:126–143.

U.S. Census Bureau QuickFacts: San Luis Obispo County, California; Tulare County, California; Fresno County, California; Santa Barbara County, California; California. https://www.census.gov/quickfacts/fact/table/sanluisobispocountycalifornia,tularecountycalifornia,fresnocountycalifornia,santabarbaracountycalifornia,CA/PST045221. Accessed 28 Apr. 2022.

Union of Concerned Scientists, K. Dahl, and R. Licker. 2021. Too Hot to Work: Assessing the Threats Climate Change Poses to Outdoor Workers. Union of Concerned Scientists.

US Department of Commerce, N. (n.d.). Heat Safety Tips and Resources. https://www.weather.gov/safety/heat.

US EPA, O. 2013, February 22. Summary of the Clean Water Act. Overviews and Factsheets. https://www.epa.gov/laws-regulations/summary-clean-water-act.

US EPA, O. 2015, March 2. Memorandum: Clean Water Act Section 404 Regulatory Program and Agricultural Activities. Other Policies and Guidance. https://www.epa.gov/cwa-404/memorandum-clean-water-act-section-404-regulatory-program-a nd-agricultural-activities.

Ward, L. S. 2010. Farmworkers at Risk: The Costs of Family Separation. Journal of Immigrant and Minority Health 12:672–677.

Whitman, S., G. Good, E. R. Donoghue, N. Benbow, W. Shou, and S. Mou. 1997. Mortality in Chicago attributed to the July 1995 heat wave. American Journal of Public Health 87:1515–1518.

Yeager, J. (n.d.). Extreme heat is fatal to farmworkers in the San Joaquin Valley. Can regulations keep pace? https://centerforhealthjournalism.org/2020/03/11/extreme-heat-fatal-farmworkers-san-joaquin-valley-can-regulations-keep-pace.

Zhao, Y., Z. Huang, S. Wang, J. Hu, J. Xiao, X. Li, T. Liu, W. Zeng, L. Guo, Q. Du, and W. Ma. 2019a. Morbidity burden of respiratory diseases attributable to ambient temperature: a case study in a subtropical city in China. Environmental Health 18:89.

Zhao, Y., Z. Huang, S. Wang, J. Hu, J. Xiao, X. Li, T. Liu, W. Zeng, L. Guo, Q. Du, and W. Ma. 2019b. Morbidity burden of respiratory diseases attributable to ambient temperature: a case study in a subtropical city in China. Environmental Health 18:89.

Zuo, J., S. Pullen, J. Palmer, H. Bennetts, N. Chileshe, and T. Ma. 2015. Impacts of heat waves and corresponding measures: a review. Journal of Cleaner Production 92:1–12.