ERHS 535 Final Project: Prediction of West Niles Virus Human Cases in Southern California by Precipitation and Temperature from 2006-2010

Devin, Anastasia, Grant, Ana

11/7/2018

1. Packages used in the final project

```
library(dplyr)
library(stringr)
library(tidyselect)
library(ggplot2)
library(rmarkdown)
library(RColorBrewer)
library(knitr)
library(forcats)
library(readr)
library(tidyr)
library(broom)
library(purrr)
library(scales)
library(lubridate)
library(viridis)
library(RColorBrewer)
library(scales)
library(tidyverse)
library(tigris)
library(sf)
library(cowplot)
library(gridExtra)
```

2. Datasets description

Different data sets were collected from CHHS and CDC Wonder. The report of human cases for California had data arrange by week and county for the years 2006 to 2018. The data is available in:

CHHS

Daily precipitation and temperature in California were downloaded from CDC Wonder website. Every year was pulled independently. The request form might be found in the following website.

CDC Wonder

```
report <-  read_csv( "../Data/wnv_human_cases.csv")
temp_2006 <- read_csv("../Data/temp/temp_2006.csv")
temp_2007 <- read_csv("../Data/temp/temp_2007.csv")
temp_2008 <- read_csv("../Data/temp/temp_2008.csv")
temp_2009 <- read_csv("../Data/temp/temp_2009.csv")
temp_2010 <- read_csv("../Data/temp/temp_2010.csv")
precip_2006 <- read_csv("../Data/precip/precip_2006.csv")
```

```
precip_2007 <- read_csv("../Data/precip/precip_2007.csv")
precip_2008 <- read_csv("../Data/precip/precip_2008.csv")
precip_2009 <- read_csv("../Data/precip/precip_2009.csv")
precip_2010 <- read_csv("../Data/precip/precip_2010.csv")
```

3. Tyding and organizing the original data sets

3.1. Human cases dataset

The main objective was change the *weekly* report to a *year* report for the number of cases by county

```
cases <- report %>%
  select(Year,
         county = County,
         positive_cases = "Positive Cases") %>%
  mutate(date = Year) %>%
  filter(date %in% c("2006", "2007", "2008", "2009", "2010")) %>%
  group_by(date, county) %>%
  summarize(positive_cases = sum(positive_cases)) %>%
  ungroup %>%
  arrange(date)

kable(head(cases))
```

date	county	positive_cases
2006	Alameda	1
2006	Butte	31
2006	Colusa	3
2006	Contra Costa	8
2006	El Dorado	2
2006	Fresno	11

```
kable(tail(cases))
```

date	county	positive_cases
2010	Sacramento	12
2010	San Bernardino	5
2010	San Francisco	1
2010	San Joaquin	6
2010	Stanislaus	12
2010	Tulare	12
#3.2. T	emperature datase	t

The main objective was join all the temperature data in one data set and select the relevant variables.

```
temp <- rbind(temp_2006, temp_2007, temp_2008, temp_2009, temp_2010)

temp <- temp %>%
  select(County, "County Code", "Month Day, Year Code", "Day of Year",
         "Avg Daily Max Air Temperature (F)",
         "Avg Daily Min Air Temperature (F)") %>%
```

```
  rename(county = County,
         fips = "County Code",
         date = "Month Day, Year Code",
         day_year = "Day of Year",
         max_temp = "Avg Daily Max Air Temperature (F)",
         min_temp_f = "Avg Daily Min Air Temperature (F)") %>%
  mutate(date = mdy(date))
kable(head(temp))
```

county	fips	date	day_year	max_temp	min_temp_f
Alameda County, CA	06001	2006-01-01	1	54.56	48.00
Alameda County, CA	06001	2006-01-02	2	51.49	46.46
Alameda County, CA	06001	2006-01-03	3	57.01	45.33
Alameda County, CA	06001	2006-01-04	4	62.25	49.38
Alameda County, CA	06001	2006-01-05	5	65.99	50.21
Alameda County, CA	06001	2006-01-06	6	67.36	51.97

3.3. Precipitation dataset

Same as the previous objective, precipitation datasets were joined and the relevant information was selected.

```
precip <- rbind(precip_2006, precip_2007, precip_2008, precip_2009, precip_2010)

precip <- precip %>%
  select(County, "Month Day, Year Code", "Avg Daily Precipitation (mm)")%>%
  rename(county = County,
         date = "Month Day, Year Code",
         avg_precip = "Avg Daily Precipitation (mm)") %>%
  mutate(date = mdy(date))

kable(head(precip))
```

county	date	avg_precip
Alameda County, CA	2006-01-01	3
Alameda County, CA	2006-01-02	31
Alameda County, CA	2006-01-03	1
Alameda County, CA	2006-01-04	0
Alameda County, CA	2006-01-05	0
Alameda County, CA	2006-01-06	0

4. Merging cleaned datasets

Two approaches were considered.

- First, precipitation and temperature were merged and year averages for this two variables were calculated.
- Second,the yearly human West Niles Virus cases dataset was joined with weather organize by year.

4.1. Weather data

In this step, temperature and precipitation data sets were merged.

```
ca_weather <- merge(temp, precip, by = c("county", "date")) %>%
  select(county, date, fips, avg_precip, max_temp) %>%
  separate(county, c("county", "state"), sep = " County, CA") %>%
  select(county, date, fips, avg_precip, max_temp) %>%
  mutate(date = year(date)) %>%
  group_by(county, fips, date) %>%
  summarise(avg_precip = mean(avg_precip), avg_max_temp = mean(max_temp)) %>%
  ungroup

kable(tail(ca_weather))
```

county	fips	date	avg_precip	avg_max_temp
Yolo	06113	2010	1.747945	73.43753
Yuba	06115	2006	2.838356	71.49586
Yuba	06115	2007	1.868493	72.56929
Yuba	06115	2008	1.792350	72.60995
Yuba	06115	2009	2.238356	72.58660
Yuba	06115	2010	3.378082	69.89458

4.2. Cases and weather data

In this step cases and weather dataset were joined

```
ca_weather_cases <- full_join(ca_weather, cases, by = c('date', 'county')) %>%
  mutate(positive_cases = ifelse(!is.na(positive_cases), positive_cases, 0)) %>%
  arrange(positive_cases)

kable(head(ca_weather_cases))
```

county	fips	date	avg_precip	avg_max_temp	positive_cases
Alameda	06001	2007	0.7506849	71.00534	0
Alameda	06001	2009	1.0739726	71.20956	0
Alpine	06003	2006	2.6575342	50.47945	0
Alpine	06003	2007	1.6082192	51.38460	0
Alpine	06003	2008	2.0327869	52.24068	0
Alpine	06003	2009	2.3671233	51.35307	0

5. Creating the final data set

To create the final dataset and in order to be able to create a Shinny App, geographical data from California counties was joined with human cases and weather data (joined in part 4.2).

```
kable(head(ca_county_cases, 1))
```

STATEFP	COUNTYFP	COUNTYNS	AFFGEOID	GEOID	NAME	LSAD	ALAND
06	037	00277283	0500000US06037	06037	Los Angeles	06	10510651024
-118.38203	7, -118.377	169, -118.3	70323, -118.36830	1, -118.	36727, -118.37	0161, -	118.368126, -1
-118.36780	6, -118.374	768, -118.3	84564, -118.39305	5, -118.	402941, -118.4	08543,	-118.413767, -
-118.51169	6, -118.513	36, -118.51	4073, -118.515914	, -118.5	16267, -118.52	323, -1	18.533681, -11
33.466603,	33.466355,	33.462016,	33.455313, 33.45	3302, 33	.449592, 33.44	5931, 3	3.443352, 33.4

STATEFP	COUNTYFP	COUNTYNS	AFFGEOID	GEOID	NAME	LSAD	ALAND
33.3485284	995781, 33.	344728, 33.	3442993943942, 33	.342695,	33.3408250322	726, 33	.334781, 33.33
33.383238,	33.385818,	33.38632,	33.388994, 33.396	108, 33.	401176, 33.404	67, 33.	409233, 33.412
list(c	(-118.6105,	-118.59885	3, -118.595763, -	118.5938	62, -118.58673	2, -118	.57988, -118.5
-118.4	06095, -118	.40871, -11	8.41584, -118.422	733, -11	8.427249, -118	.431809	, -118.434141,
-118.5	61347, -118	.56439, -11	8.574087, -118.58	0362, -1	18.582263, -11	8.58188	3, -118.584735
32.848	233, 32.839	047, 32.820	919, 32.817078, 3	2.820873	, 32.822471, 3	2.82494	8, 32.826266,
32.934	252, 32.947	018, 32.951	805, 32.958506, 3	2.964887	, 32.974299, 3	2.98259	4, 32.987698,
-118.7	03392, -118	.693834, -1	18.67566, -118.67	4241, -1	18.668152, -11	8.66815	3, -118.668162
-118.7	38618, -118	.796842, -1	18.821077, -118.8	81364, -	118.887362, -1	18.8944	74, -118.88438
-118.0	23406, -117	.774368, -1	17.667292, -117.6	67244, -	117.667108, -1	17.6673	23, -117.66708
-117.6	77576, -117	.681713, -1	17.683264, -117.6	85551, -	117.686799, -1	17.6891	71, -117.69354
-117.7	35344, -117	.735846, -1	17.744066, -117.7	45378, -	117.745417, -1	17.7676	9, -117.767752
-117.9	25836, -117	.927483, -1	17.928783, -117.9	28658, -	117.929148, -1	17.9316	23, -117.93969
-117.9	8535, -117.	985342, -11	7.9877, -117.9896	83, -117	.994136, -117.	994119,	-117.994044,
-118.0	6325, -118.	063162, -11	8.067639, -118.06	8824, -1	18.069908, -11	8.07182	7, -118.07184,
-118.1	59907459484	, -118.1664	97421716, -118.16	6887, -1	18.1755, -118.	1761987	47974, -118.17
-118.2	8589791404,	-118.28799	8876897, -118.288	373, -11	8.292873356242	, -118.	294788, -118.2
-118.3	8979429006,	-118.39289	5696474, -118.398	286, -11	8.402885, -118	.404415	, -118.411211,
-118.4	06999, -118	.4074085689	28, -118.408297,	-118.405	007, -118.4000	2228594	9, -118.394376
-118.4	12695385473	, -118.4127	08, -118.41465406	7515, -1	18.41561099890	9, -118	.416260078565,
-118.4	8755203382,	-118.48758	4492053, -118.493	11546350	2, -118.495124	882367,	-118.502813,
-118.6	22757, -118	.6234934343	14, -118.626946,	-118.635	250350277, -11	8.63592	0764079, -118.
-118.7	87095766528	, -118.7871	38163409, -118.79	19155422	34, -118.79333	1, -118	.794289, -118.
-118.9	44502, 34.0	46563, 34.0	74967, 34.104817,	34.1142	1, 34.11964, 3	4.12676	5, 34.140882,
34.240	426, 34.241	948, 34.254	93, 34.263351, 34	.268322,	34.269522, 34	.269812	, 34.270242, 3
34.818	025, 34.818	303, 34.819	726, 34.820103, 3	4.820146	, 34.820138, 3	4.82051	3, 34.820568,
34.232	393, 34.229	76, 34.2203	34, 34.166103, 34	.165173,	34.164368, 34	.155827	, 34.150488, 3
34.019	808, 34.020	018, 34.020	024, 34.023506, 3	4.019429	, 34.013567, 3	4.01119	2, 34.004611,
33.946	076, 33.946	072, 33.946	069, 33.946067, 3	3.946066	, 33.946054, 3	3.94605	8, 33.946011,
33.873	322, 33.866	242, 33.866	013, 33.862369, 3	3.862332	, 33.858723, 3	3.85876	8, 33.8578, 33
33.747	09, 33.7469	85, 33.7451	76, 33.7432945549	888, 33.	7446677877779,	33.753	1579426909, 33
33.710	7504081902,	33.7097317	630176, 33.705024	, 33.704	03, 33.7048765	794076,	33.707091, 33
33.738	086, 33.736	315, 33.736	4102442351, 33.73	73192380	487, 33.737717	7991397	, 33.737934, 3
33.782	015, 33.782	7560765859,	33.7851490520678	, 33.785	2926587749, 33	.787919	, 33.79189, 33
33.876	7032178307,	33.8769113	901839, 33.883884	2161597,	33.883913, 33	.887608	2481903, 33.88
33.999	8102929439,	33.9998436	821985, 34.005533	3023367,	34.0076003587	566, 34	.015509, 34.01
34.036	3015525377,	34.0362558	229882, 34.036229	, 34.036	262975196, 34.	03793,	34.037531, 34.
34.001	239, 34.000	198, 34.001	622, 34.001829050	1336, 34	.0138039038131	, 34.01	39390210272, 3

```
unique(ca_county_cases$county)
```

```
##  [1] "Los Angeles"     "Santa Barbara"   "San Bernardino"
##  [4] "Ventura"         "Orange"          "Imperial"
##  [7] "San Luis Obispo" "Riverside"       "Kern"
## [10] "San Diego"
```

```
unique(ca_county_cases$date)
```

```
## [1] 2010 2006 2009 2007 2008
```