Final_Project_Markdown

Human Epi 2

11/7/2018

Packages used in the final project

```r
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)
```

Uploading California weather datasets

Datasets were downloaded from CDC Wonder for the years 2006 to 2011 (from January 1st to December 31st)

CDC Wonder

West Niles Data:

```r
report <-  read_csv( "../Data/wnv_human_cases.csv")
```

Change week to date

```r
a <- report %>%
  filter(Year == "2006")
a <- a %>%
  mutate(Date = lubridate::ymd( "2006-01-01" ) + lubridate::weeks(a$`Week Reported` - 1 ))
b <- report %>%
  filter(Year == "2007")
b <- b %>%
  mutate(Date = lubridate::ymd( "2007-01-01" ) + lubridate::weeks(b$`Week Reported` - 1 ))
c <- report %>%
  filter(Year == "2008")
c <- c %>%
```

```r
  mutate(Date = lubridate::ymd( "2008-01-01" ) + lubridate::weeks(c$`Week Reported` - 1 ))
d <- report %>%
  filter(Year == "2009")
d <- d %>%
  mutate(Date = lubridate::ymd( "2009-01-01" ) + lubridate::weeks(d$`Week Reported` - 1 ))
e <- report %>%
  filter(Year == "2010")
e <- e %>%
  mutate(Date = lubridate::ymd( "2010-01-01" ) + lubridate::weeks(e$`Week Reported` - 1 ))

ab <- merge(a,b, all = TRUE)
abc <- merge(ab, c, all = TRUE)
abcd <- merge(abc, d, all = TRUE)
abcde <- merge(abcd, e, all = TRUE)
```

```r
cases <- abcde %>%
  mutate(month = month(Date)) %>%
  select(year = Year,
         county = County,
         positive_cases = "Positive Cases",
         month) %>%
  group_by(year, county, month) %>%
  summarize(positive_cases = sum(positive_cases)) %>%
  ungroup

cases
```

```
## # A tibble: 334 x 4
##     year county        month positive_cases
##    <int> <chr>         <dbl>          <int>
##  1  2006 Alameda           8              1
##  2  2006 Butte             8             15
##  3  2006 Butte             9             15
##  4  2006 Butte            10              1
##  5  2006 Colusa            7              1
##  6  2006 Colusa            8              1
##  7  2006 Colusa           10              1
##  8  2006 Contra Costa      8              4
##  9  2006 Contra Costa      9              4
## 10  2006 El Dorado         8              1
## # ... with 324 more rows
```

Temperature and precipitation data:

1. Uploading datasets

```r
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")
temp_2011 <- read_csv("../Data/temp/temp_2011.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")
precip_2011 <- read_csv("../Data/precip/precip_2011.csv")
```

2. Creating a dataframe with air temperature (°F) information of California for the years 2006 to 2011.

```
temp <- rbind(temp_2006, temp_2007, temp_2008, temp_2009,
              temp_2010, temp_2011)

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

```
## # A tibble: 6 x 6
##   county            fip   date       day_year max_temp_f min_temp_f
##   <chr>             <chr> <date>        <int>      <dbl>      <dbl>
## 1 Alameda County, CA 06001 2006-01-01        1       54.6        48
## 2 Alameda County, CA 06001 2006-01-02        2       51.5       46.5
## 3 Alameda County, CA 06001 2006-01-03        3       57.0       45.3
## 4 Alameda County, CA 06001 2006-01-04        4       62.2       49.4
## 5 Alameda County, CA 06001 2006-01-05        5       66.0       50.2
## 6 Alameda County, CA 06001 2006-01-06        6       67.4       52.0
```

```
tail(temp)
```

```
## # A tibble: 6 x 6
##   county          fip   date       day_year max_temp_f min_temp_f
##   <chr>           <chr> <date>        <int>      <dbl>      <dbl>
## 1 Yuba County, CA 06115 2011-12-26      360       52.2       33.5
## 2 Yuba County, CA 06115 2011-12-27      361       56.5       34.0
## 3 Yuba County, CA 06115 2011-12-28      362       56.0       44.7
## 4 Yuba County, CA 06115 2011-12-29      363       61.8       46.7
## 5 Yuba County, CA 06115 2011-12-30      364       56.4       44.7
## 6 Yuba County, CA 06115 2011-12-31      365       58.5       42.1
```

3. Creating a dataframe with the precipitation (mm) information of California for the years 2006 to 2011.

```
precip <- rbind(precip_2006, precip_2007, precip_2008, precip_2009,
                precip_2010, precip_2011)

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))
```

```
head(precip)
```

```
## # A tibble: 6 x 3
##   county            date       avg_precip
##   <chr>             <date>          <int>
## 1 Alameda County, CA 2006-01-01          3
## 2 Alameda County, CA 2006-01-02         31
## 3 Alameda County, CA 2006-01-03          1
## 4 Alameda County, CA 2006-01-04          0
## 5 Alameda County, CA 2006-01-05          0
## 6 Alameda County, CA 2006-01-06          0
```

```
tail(precip)
```

```
## # A tibble: 6 x 3
##   county         date       avg_precip
##   <chr>          <date>          <int>
## 1 Yuba County, CA 2011-12-26          0
## 2 Yuba County, CA 2011-12-27          0
## 3 Yuba County, CA 2011-12-28          0
## 4 Yuba County, CA 2011-12-29          0
## 5 Yuba County, CA 2011-12-30          1
## 6 Yuba County, CA 2011-12-31          0
```

4. Merging temperature and precipitation of CA by date and county to have just one main dataframe
   with weather conditions for the years 2006 to 2011.

```
ca_weather <- merge(temp, precip, by = c("county", "date"))
head(ca_weather)
```

```
##                county       date    fip day_year max_temp_f min_temp_f
## 1 Alameda County, CA 2006-01-01 06001        1      54.56      48.00
## 2 Alameda County, CA 2006-01-02 06001        2      51.49      46.46
## 3 Alameda County, CA 2006-01-03 06001        3      57.01      45.33
## 4 Alameda County, CA 2006-01-04 06001        4      62.25      49.38
## 5 Alameda County, CA 2006-01-05 06001        5      65.99      50.21
## 6 Alameda County, CA 2006-01-06 06001        6      67.36      51.97
##   avg_precip
## 1          3
## 2         31
## 3          1
## 4          0
## 5          0
## 6          0
```

```
tail(ca_weather)
```

```
##                     county       date    fip day_year max_temp_f min_temp_f
## 127073 Yuba County, CA 2011-12-26 06115      360      52.21      33.46
## 127074 Yuba County, CA 2011-12-27 06115      361      56.54      34.04
## 127075 Yuba County, CA 2011-12-28 06115      362      56.05      44.69
## 127076 Yuba County, CA 2011-12-29 06115      363      61.79      46.68
## 127077 Yuba County, CA 2011-12-30 06115      364      56.38      44.66
## 127078 Yuba County, CA 2011-12-31 06115      365      58.51      42.12
```

```
##         avg_precip
## 127073          0
## 127074          0
## 127075          0
## 127076          0
## 127077          1
## 127078          0
```

```r
ca_precip <- ca_weather %>%
  select(county, date, fip, avg_precip) %>%
  separate(county, c("county", "state"), sep = " County, CA") %>%
  select(county, date, fip, avg_precip) %>%
  mutate(month = month(date)) %>%
  mutate(year = year(date)) %>%
  group_by(county, fip, month, year) %>%
  summarise(avg_precip = mean(avg_precip)) %>%
  ungroup %>%
  arrange(year)

head(ca_precip)
```

```
## # A tibble: 6 x 5
##   county  fip   month  year avg_precip
##   <chr>   <chr> <dbl> <dbl>      <dbl>
## 1 Alameda 06001     1  2006      2.19
## 2 Alameda 06001     2  2006      1.5
## 3 Alameda 06001     3  2006      4.32
## 4 Alameda 06001     4  2006      2.97
## 5 Alameda 06001     5  2006      0.355
## 6 Alameda 06001     6  2006      0
```

Merge datasets

```r
ca_precip_cases <- full_join(ca_precip, cases, by = c('month', 'year', 'county'))

ca_precip_cases$positive_cases[is.na(ca_precip_cases$positive_cases)] <- 0

ca_precip_cases <- ca_precip_cases %>%
  arrange(desc(positive_cases))

head(ca_precip_cases)
```

```
## # A tibble: 6 x 6
##   county      fip   month  year avg_precip positive_cases
##   <chr>       <chr> <dbl> <dbl>      <dbl>          <dbl>
## 1 Los Angeles 06037     9  2008     0                  55
## 2 Kern        06029     8  2007     0.0323             53
## 3 Kern        06029     7  2007     0.0323             37
## 4 Los Angeles 06037    10  2008     0.0645             35
## 5 Kern        06029     9  2007     0.167              26
## 6 Orange      06059     9  2008     0                  26
```