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)
library(sf)
library(tigris)
library(shiny)
```

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 %>%
  select(date,
         county = County,
         positive_cases = "Positive Cases") %>%
  mutate(date = ymd(date)) %>%
  mutate(date = floor_date(date, unit = "month")) %>%
  group_by(date, county) %>%
  summarize(positive_cases = sum(positive_cases)) %>%
  ungroup

cases
```

```
## # A tibble: 334 x 3
##    date       county       positive_cases
##    <date>     <chr>                 <int>
##  1 2006-06-01 Kern                      1
##  2 2006-07-01 Colusa                    1
##  3 2006-07-01 Fresno                    1
##  4 2006-07-01 Kern                     10
##  5 2006-07-01 Sacramento                1
##  6 2006-07-01 San Joaquin               1
##  7 2006-07-01 Stanislaus                1
##  8 2006-07-01 Sutter                    1
##  9 2006-07-01 Yolo                      3
## 10 2006-08-01 Alameda                   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")
```

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

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

```r
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,
         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
```

```r
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 2010-12-26      360       49.6       41.2
## 2 Yuba County, CA 06115 2010-12-27      361       51.7       41.1
## 3 Yuba County, CA 06115 2010-12-28      362       46.0       41.4
## 4 Yuba County, CA 06115 2010-12-29      363       48.1       32.4
## 5 Yuba County, CA 06115 2010-12-30      364       43.2       29.0
## 6 Yuba County, CA 06115 2010-12-31      365       42.5       27.3
```

3. Creating a dataframe with the precipitation (mm) information of California for the years 2006 to 2010.

```r
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))

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 2010-12-26          7
## 2 Yuba County, CA 2010-12-27          0
## 3 Yuba County, CA 2010-12-28         32
## 4 Yuba County, CA 2010-12-29         10
## 5 Yuba County, CA 2010-12-30          0
## 6 Yuba County, CA 2010-12-31          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
## 105903 Yuba County, CA 2010-12-26 06115      360      49.55      41.24
## 105904 Yuba County, CA 2010-12-27 06115      361      51.68      41.14
## 105905 Yuba County, CA 2010-12-28 06115      362      46.04      41.42
## 105906 Yuba County, CA 2010-12-29 06115      363      48.10      32.41
## 105907 Yuba County, CA 2010-12-30 06115      364      43.17      29.02
## 105908 Yuba County, CA 2010-12-31 06115      365      42.46      27.32
##        avg_precip
## 105903          7
## 105904          0
## 105905         32
## 105906         10
```

```
## 105907          0
## 105908          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(date = floor_date(date, unit = "month")) %>%
  group_by(county, fip, date) %>%
  summarise(avg_precip = mean(avg_precip)) %>%
  ungroup %>%
  arrange(date)

head(ca_precip)
```

```
## # A tibble: 6 x 4
##   county    fip   date       avg_precip
##   <chr>     <chr> <date>          <dbl>
## 1 Alameda   06001 2006-01-01       2.19
## 2 Alpine    06003 2006-01-01       5.39
## 3 Amador    06005 2006-01-01       4.90
## 4 Butte     06007 2006-01-01       6.52
## 5 Calaveras 06009 2006-01-01       5.48
## 6 Colusa    06011 2006-01-01       2.06
```

#Merge datasets

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

ca_precip_cases <- ca_precip_cases %>%
  mutate(positive_cases = ifelse(!is.na(positive_cases), positive_cases, 0))

ca_precip_cases <- ca_precip_cases %>%
  arrange(positive_cases) %>%
  rename(fips = fip)

head(ca_precip_cases)
```

```
## # A tibble: 6 x 5
##   county    fips  date       avg_precip positive_cases
##   <chr>     <chr> <date>          <dbl>          <dbl>
## 1 Alameda   06001 2006-01-01       2.19              0
## 2 Alpine    06003 2006-01-01       5.39              0
## 3 Amador    06005 2006-01-01       4.90              0
## 4 Butte     06007 2006-01-01       6.52              0
## 5 Calaveras 06009 2006-01-01       5.48              0
## 6 Colusa    06011 2006-01-01       2.06              0
```