Reformatting Phoenix, AZ Weather Data

```
#load packages

library(readr)
library(plyr)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:plyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
library(tsibble)
```

```
##
## Attaching package: 'tsibble'

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, union
```

```
raw_data <- read_csv(file= 'data/weather_ontario_peddleton_raw.csv')
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   STATION = col_character(),
##   NAME = col_character(),
##   DATE = col_date(format = ""),
##   PRCP = col_double(),
##   SNOW = col_double(),
##   SNWD = col_double(),
##   TAVG = col_double(),
##   TMAX = col_double(),
##   TMIN = col_double(),
##   TOBS = col_logical()
## )
```

```
## Warning: 703 parsing failures.
##  row  col           expected actual                                  file
## 1659 TOBS 1/0/T/F/TRUE/FALSE     79 'data/weather_ontario_peddleton_raw.csv'
## 1660 TOBS 1/0/T/F/TRUE/FALSE     85 'data/weather_ontario_peddleton_raw.csv'
## 1661 TOBS 1/0/T/F/TRUE/FALSE     78 'data/weather_ontario_peddleton_raw.csv'
## 1662 TOBS 1/0/T/F/TRUE/FALSE     79 'data/weather_ontario_peddleton_raw.csv'
## 1663 TOBS 1/0/T/F/TRUE/FALSE     79 'data/weather_ontario_peddleton_raw.csv'
## .... .... .................. ...... ....................................
## See problems(...) for more details.
```

```
raw_data
```

```
## # A tibble: 4,019 x 10
##    STATION   NAME           DATE        PRCP  SNOW  SNWD  TAVG  TMAX  TMIN TOBS
##    <chr>     <chr>          <date>     <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <lgl>
##  1 USW00024~ PENDLETON E O~ 2017-01-01  0.07   0.9   1.2    35    37    26 NA
##  2 USW00024~ PENDLETON E O~ 2017-01-02  0.04   0.8   2      26    26    21 NA
##  3 USW00024~ PENDLETON E O~ 2017-01-03  0      0     1.2    20    21    16 NA
##  4 USW00024~ PENDLETON E O~ 2017-01-04  0      0.1   1.2    15    16     2 NA
##  5 USW00024~ PENDLETON E O~ 2017-01-05  0      0     1.2     6    16     0 NA
##  6 USW00024~ PENDLETON E O~ 2017-01-06  0      0     1.2     8    18     0 NA
##  7 USW00024~ PENDLETON E O~ 2017-01-07  0.11   1.4   2       6    15    -1 NA
##  8 USW00024~ PENDLETON E O~ 2017-01-08  0.2    0.3   1.2    15    21    15 NA
##  9 USW00024~ PENDLETON E O~ 2017-01-09  0.13   1.5   3.1    26    41    21 NA
## 10 USW00024~ PENDLETON E O~ 2017-01-10  0.38   4.5   7.1    30    36    17 NA
## # ... with 4,009 more rows
```

```
tail(raw_data)
```

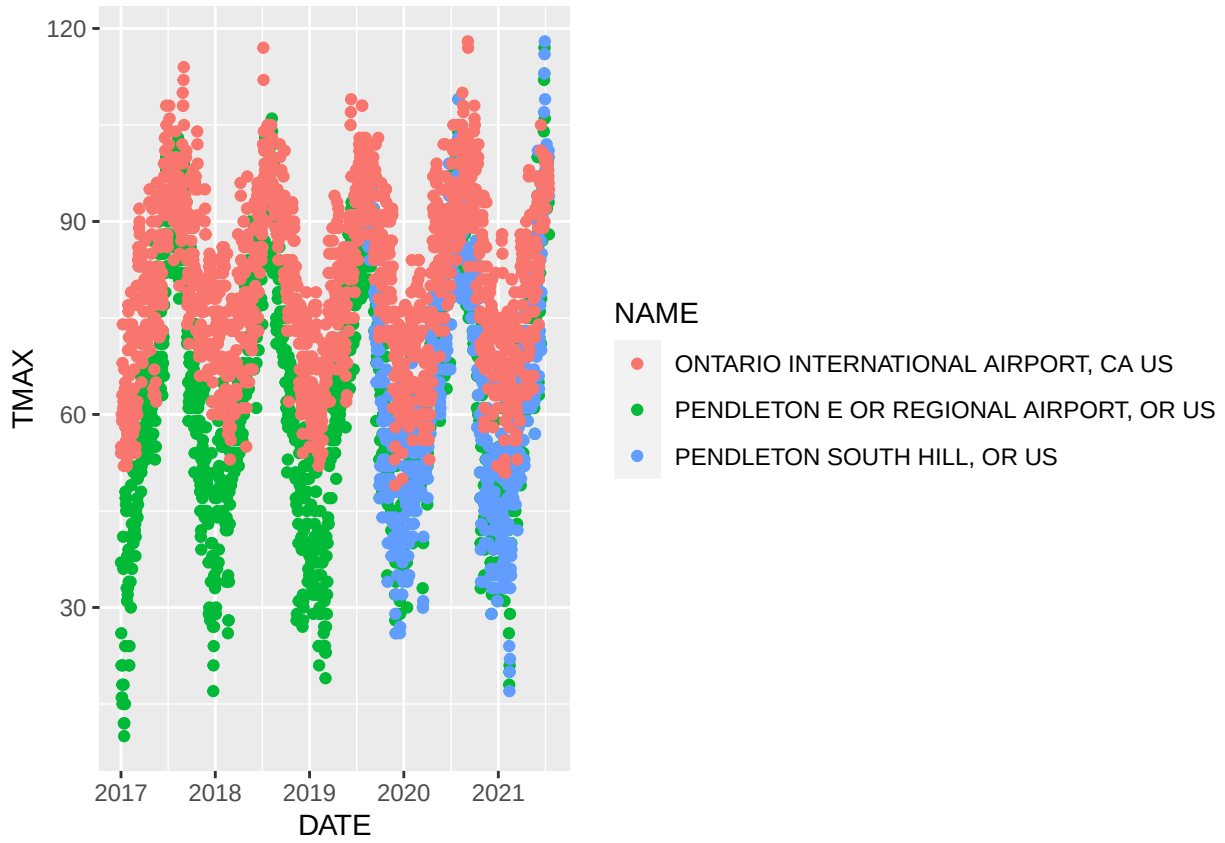
```
## # A tibble: 6 x 10
##   STATION   NAME            DATE        PRCP  SNOW  SNWD  TAVG  TMAX  TMIN TOBS
##   <chr>     <chr>           <date>     <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <lgl>
## 1 USW00003~ ONTARIO INTERN~ 2021-07-11  0       NA    NA    NA    99    70 NA
## 2 USW00003~ ONTARIO INTERN~ 2021-07-12  0       NA    NA    NA    94    67 NA
## 3 USW00003~ ONTARIO INTERN~ 2021-07-13  0       NA    NA    NA    98    70 NA
## 4 USW00003~ ONTARIO INTERN~ 2021-07-14  0.05    NA    NA    NA    97    71 NA
## 5 USW00003~ ONTARIO INTERN~ 2021-07-15  0       NA    NA    NA    96    68 NA
## 6 USW00003~ ONTARIO INTERN~ 2021-07-16  0       NA    NA    NA    95    66 NA
```

```
colnames(raw_data)
```

```
##  [1] "STATION" "NAME"    "DATE"    "PRCP"    "SNOW"    "SNWD"    "TAVG"
##  [8] "TMAX"    "TMIN"    "TOBS"
```

```
#plot
ggplot(raw_data, aes(x = DATE, y = TMAX, color = NAME)) +
  geom_point()
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
```




```
station_data <- raw_data %>%
  filter(STATION == "USW00024155") %>%
  select(DATE, PRCP, TMAX)

head(station_data)
```

```
## # A tibble: 6 x 3
##   DATE        PRCP  TMAX
##   <date>     <dbl> <dbl>
## 1 2017-01-01  0.07    37
## 2 2017-01-02  0.04    26
## 3 2017-01-03  0       21
## 4 2017-01-04  0       16
## 5 2017-01-05  0       16
## 6 2017-01-06  0       18
```

```
dim(station_data)
```

```
## [1] 1658    3
```

```
# create a yearweek column
data_yw <- station_data %>%
  mutate(yw = yearweek(DATE))

data_yw
```

```
## # A tibble: 1,658 x 4
##    DATE        PRCP  TMAX       yw
##    <date>     <dbl> <dbl>   <week>
##  1 2017-01-01  0.07    37 2016 W52
##  2 2017-01-02  0.04    26 2017 W01
##  3 2017-01-03  0       21 2017 W01
##  4 2017-01-04  0       16 2017 W01
##  5 2017-01-05  0       16 2017 W01
##  6 2017-01-06  0       18 2017 W01
##  7 2017-01-07  0.11    15 2017 W01
##  8 2017-01-08  0.2     21 2017 W01
##  9 2017-01-09  0.13    41 2017 W02
## 10 2017-01-10  0.38    36 2017 W02
## # ... with 1,648 more rows
```

```
aggregate_weekly <- data_yw %>%
  ddply(. (yw), numcolwise(mean, na.rm=TRUE)) %>%
  as_tsibble(index = yw)

aggregate_weekly
```

```
## # A tsibble: 238 x 3 [1W]
##          yw    PRCP  TMAX
##      <week>   <dbl> <dbl>
##  1 2016 W52 0.07     37
##  2 2017 W01 0.05     19
##  3 2017 W02 0.0914   20.6
##  4 2017 W03 0.0386   38.9
##  5 2017 W04 0.00286  34.6
##  6 2017 W05 0.0571   33.1
##  7 2017 W06 0.0343   43.4
##  8 2017 W07 0.126    48.1
##  9 2017 W08 0.12     44.7
## 10 2017 W09 0.0786   49.1
## # ... with 228 more rows
```

```
write.csv(aggregate_weekly, "data/OR_PEN_weather.csv")
```