ANA515- Assignment 2 - Week 4

Sandeep Tom

11/14/2021

#Section 1: Description of the data. The data we are using are temperature measurements that were taken across different cities in the United stats over the course of a 12-month period June 2014 to June 2015. The data taken across multiple cities tells us what the year looked like with record setting temperatures. The data used here is the for the city of Philadelphia.

#Section 3: Clean the data: # Adding a filter to the datframe weather\_data

weather\_PHL <-filter(weather\_data,actual\_mean\_temp >= 19)

# renamed the column date to date\_of\_measurement

rename(weather\_PHL,date\_of\_measurement = date)

## # A tibble: 358 x 13  
## date\_of\_measurement actual\_mean\_temp actual\_min\_temp actual\_max\_temp  
## <chr> <dbl> <dbl> <dbl>  
## 1 2014-7-1 83 72 93  
## 2 2014-7-2 86 75 96  
## 3 2014-7-3 83 74 92  
## 4 2014-7-4 73 68 78  
## 5 2014-7-5 74 64 83  
## 6 2014-7-6 75 64 86  
## 7 2014-7-7 84 74 93  
## 8 2014-7-8 84 72 95  
## 9 2014-7-9 80 71 89  
## 10 2014-7-10 81 74 87  
## # ... with 348 more rows, and 9 more variables: average\_min\_temp <dbl>,  
## # average\_max\_temp <dbl>, record\_min\_temp <dbl>, record\_max\_temp <dbl>,  
## # record\_min\_temp\_year <dbl>, record\_max\_temp\_year <dbl>,  
## # actual\_precipitation <dbl>, average\_precipitation <dbl>,  
## # record\_precipitation <dbl>

# Selecting 4 coulumns from the dataframe

weather\_PHL %>%  
 select(1,2,3,4)

## # A tibble: 358 x 4  
## date actual\_mean\_temp actual\_min\_temp actual\_max\_temp  
## <chr> <dbl> <dbl> <dbl>  
## 1 2014-7-1 83 72 93  
## 2 2014-7-2 86 75 96  
## 3 2014-7-3 83 74 92  
## 4 2014-7-4 73 68 78  
## 5 2014-7-5 74 64 83  
## 6 2014-7-6 75 64 86  
## 7 2014-7-7 84 74 93  
## 8 2014-7-8 84 72 95  
## 9 2014-7-9 80 71 89  
## 10 2014-7-10 81 74 87  
## # ... with 348 more rows

#Section 4: Characteristics of the data We have data about 358 rows and 13columns. The data is temperature measurements that were taken in the city of Philadelphia over the course of a 12 month period. The columns include the data that the temperature was recorded and the actual mean, min and max temperature.

# Creating a table by seclecting 4 columns.

weather\_PHL\_TABLE <- weather\_PHL %>%  
 select(1,2,3,4)  
weather\_PHL\_TABLE = apply\_labels(weather\_PHL, date = "date the temeprature was measured", actual\_mean\_temp ="actual mean temperature on the date" , actual\_min\_temp ="Actual minimum temeprature measured on the date", actual\_max\_temp = "Actual maximum temeperature measured on the date")  
print(weather\_PHL\_TABLE)

## # A tibble: 358 x 13  
## date actual\_mean\_temp actual\_min\_temp actual\_max\_temp average\_min\_temp  
## <labelled> <labelled> <labelled> <labelled> <dbl>  
## 1 2014-7-1 83 72 93 68  
## 2 2014-7-2 86 75 96 68  
## 3 2014-7-3 83 74 92 68  
## 4 2014-7-4 73 68 78 68  
## 5 2014-7-5 74 64 83 69  
## 6 2014-7-6 75 64 86 69  
## 7 2014-7-7 84 74 93 69  
## 8 2014-7-8 84 72 95 69  
## 9 2014-7-9 80 71 89 69  
## 10 2014-7-10 81 74 87 69  
## # ... with 348 more rows, and 8 more variables: average\_max\_temp <dbl>,  
## # record\_min\_temp <dbl>, record\_max\_temp <dbl>, record\_min\_temp\_year <dbl>,  
## # record\_max\_temp\_year <dbl>, actual\_precipitation <dbl>,  
## # average\_precipitation <dbl>, record\_precipitation <dbl>

#Section 5: Summary statistics - Selecting 3 columns from the dataframe. The columns that were used were the actual\_mean\_temp, actual\_min\_temp & actual\_max\_temp.

weather\_PHL\_SUM <- weather\_data%>%  
 select(2,3,4)  
summary(weather\_PHL\_SUM)

## actual\_mean\_temp actual\_min\_temp actual\_max\_temp  
## Min. :10.00 Min. : 2.00 Min. :17.00   
## 1st Qu.:39.00 1st Qu.:33.00 1st Qu.:46.00   
## Median :59.00 Median :50.00 Median :69.00   
## Mean :55.88 Mean :47.27 Mean :63.98   
## 3rd Qu.:73.00 3rd Qu.:64.00 3rd Qu.:82.00   
## Max. :86.00 Max. :77.00 Max. :96.00