#INCLASS3 #JESSICA CHOE

#0.

library(tidyverse)

## Warning: package 'tidyverse' was built under R version 3.6.2

## -- Attaching packages ----------------------- tidyverse 1.3.0 --

## v ggplot2 3.2.1 v purrr 0.3.3  
## v tibble 2.1.3 v dplyr 0.8.3  
## v tidyr 1.0.2 v stringr 1.4.0  
## v readr 1.3.1 v forcats 0.4.0

## Warning: package 'tidyr' was built under R version 3.6.2

## Warning: package 'readr' was built under R version 3.6.2

## Warning: package 'purrr' was built under R version 3.6.2

## Warning: package 'forcats' was built under R version 3.6.2

## -- Conflicts -------------------------- tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(dplyr)  
  
oh\_brfss <- read\_csv("https://raw.githubusercontent.com/qyuan20/Datasets-DataVis/master/ohioBRFSS2018.csv")

## Warning: Missing column names filled in: 'X1' [1]

## Parsed with column specification:  
## cols(  
## .default = col\_double(),  
## IDATE = col\_character(),  
## IMONTH = col\_character(),  
## IDAY = col\_character(),  
## COLGHOUS = col\_logical(),  
## LADULT = col\_logical(),  
## SAFETIME = col\_logical(),  
## CTELNUM1 = col\_logical(),  
## CELLFON5 = col\_logical(),  
## CADULT = col\_logical(),  
## PVTRESD3 = col\_logical(),  
## CCLGHOUS = col\_logical(),  
## CSTATE1 = col\_logical(),  
## LANDLINE = col\_logical(),  
## HHADULT = col\_logical(),  
## PDIABTST = col\_logical(),  
## PREDIAB1 = col\_logical(),  
## INSULIN = col\_logical(),  
## BLDSUGAR = col\_logical(),  
## FEETCHK3 = col\_logical(),  
## DOCTDIAB = col\_logical()  
## # ... with 68 more columns  
## )

## See spec(...) for full column specifications.

## Warning: 12980 parsing failures.  
## row col expected actual file  
## 2106 LADULT 1/0/T/F/TRUE/FALSE 2 'https://raw.githubusercontent.com/qyuan20/Datasets-DataVis/master/ohioBRFSS2018.csv'  
## 6287 CADULT 1/0/T/F/TRUE/FALSE 2 'https://raw.githubusercontent.com/qyuan20/Datasets-DataVis/master/ohioBRFSS2018.csv'  
## 6287 CSTATE1 1/0/T/F/TRUE/FALSE 2 'https://raw.githubusercontent.com/qyuan20/Datasets-DataVis/master/ohioBRFSS2018.csv'  
## 6287 LANDLINE 1/0/T/F/TRUE/FALSE 2 'https://raw.githubusercontent.com/qyuan20/Datasets-DataVis/master/ohioBRFSS2018.csv'  
## 6288 CADULT 1/0/T/F/TRUE/FALSE 2 'https://raw.githubusercontent.com/qyuan20/Datasets-DataVis/master/ohioBRFSS2018.csv'  
## .... ........ .................. ...... .....................................................................................  
## See problems(...) for more details.

#1

oh\_selected <-select(oh\_brfss,IMONTH, IDAY, PHYSHLTH, MENTHLTH)  
head(oh\_selected)

## # A tibble: 6 x 4  
## IMONTH IDAY PHYSHLTH MENTHLTH  
## <chr> <chr> <dbl> <dbl>  
## 1 01 31 88 12  
## 2 01 29 7 88  
## 3 01 28 88 88  
## 4 01 26 1 88  
## 5 01 25 30 10  
## 6 01 26 88 88

#2

class(oh\_selected$MENTHLTH)

## [1] "numeric"

oh\_PHYSHLTH2=oh\_selected %>%   
 mutate(P.health=ifelse((PHYSHLTH==88),0,PHYSHLTH))  
  
head(oh\_PHYSHLTH2,3)

## # A tibble: 3 x 5  
## IMONTH IDAY PHYSHLTH MENTHLTH P.health  
## <chr> <chr> <dbl> <dbl> <dbl>  
## 1 01 31 88 12 0  
## 2 01 29 7 88 7  
## 3 01 28 88 88 0

Oh\_none\_removed=oh\_PHYSHLTH2 %>%   
 mutate(M.health=ifelse((MENTHLTH==88),0,MENTHLTH))  
  
colnames(Oh\_none\_removed)[1]<-"Month"   
colnames(Oh\_none\_removed)[2]<-"Day"   
  
#change column name IMONTH to Month for better visualization  
head(Oh\_none\_removed)

## # A tibble: 6 x 6  
## Month Day PHYSHLTH MENTHLTH P.health M.health  
## <chr> <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 01 31 88 12 0 12  
## 2 01 29 7 88 7 0  
## 3 01 28 88 88 0 0  
## 4 01 26 1 88 1 0  
## 5 01 25 30 10 30 10  
## 6 01 26 88 88 0 0

#3. we need PHYSHLTH from 0 to 30, same range for MENTHLTH, so anything below 31 can be filtered in this data

oh\_filtered <-Oh\_none\_removed%>%  
 dplyr::filter(P.health< 31 & M.health< 31)  
  
oh\_cleaned<-select(oh\_filtered,Month,Day,P.health,M.health)  
oh\_cleaned

## # A tibble: 12,415 x 4  
## Month Day P.health M.health  
## <chr> <chr> <dbl> <dbl>  
## 1 01 31 0 12  
## 2 01 29 7 0  
## 3 01 28 0 0  
## 4 01 26 1 0  
## 5 01 25 30 10  
## 6 01 26 0 0  
## 7 02 13 8 30  
## 8 01 31 0 0  
## 9 01 29 0 0  
## 10 01 26 1 0  
## # ... with 12,405 more rows

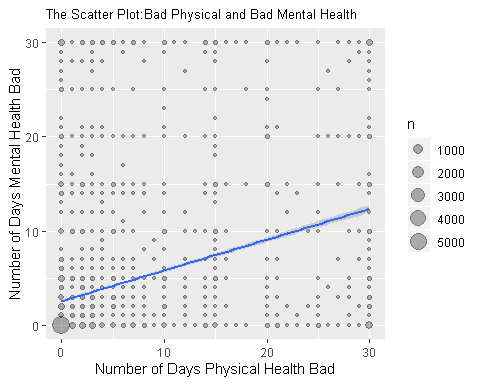
#4.

head(oh\_cleaned,6)

## # A tibble: 6 x 4  
## Month Day P.health M.health  
## <chr> <chr> <dbl> <dbl>  
## 1 01 31 0 12  
## 2 01 29 7 0  
## 3 01 28 0 0  
## 4 01 26 1 0  
## 5 01 25 30 10  
## 6 01 26 0 0

#5

ggplot(aes(x=P.health,y=M.health),data=oh\_cleaned) +  
 geom\_count(alpha=0.3)+  
 geom\_smooth( method=lm )+  
 labs(x="Number of Days Physical Health Bad",  
 y="Number of Days Mental Health Bad",   
 title="The Scatter Plot:Bad Physical and Bad Mental Health")+  
 theme(plot.title = element\_text(size = 10))

 #6.

head(oh\_cleaned,2)

## # A tibble: 2 x 4  
## Month Day P.health M.health  
## <chr> <chr> <dbl> <dbl>  
## 1 01 31 0 12  
## 2 01 29 7 0

oh\_summary=oh\_cleaned %>% group\_by(Month) %>%   
 summarise(mean.physical=mean(P.health),  
 sd.physical=sd(P.health),  
 mean.mental=mean(M.health),  
 sd.mental=sd(M.health))  
oh\_summary

## # A tibble: 12 x 5  
## Month mean.physical sd.physical mean.mental sd.mental  
## <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 01 4.71 8.89 3.44 8.00  
## 2 02 4.64 8.94 3.86 8.32  
## 3 03 5.14 9.49 3.99 8.42  
## 4 04 4.15 8.62 4.14 8.75  
## 5 05 5.25 9.67 3.90 8.38  
## 6 06 5.51 9.93 4.34 8.83  
## 7 07 4.83 9.30 4.26 8.70  
## 8 08 4.28 9.01 3.96 8.42  
## 9 09 5.06 9.50 4.33 8.76  
## 10 10 4.83 9.27 4.66 9.03  
## 11 11 4.90 9.58 4.47 9.05  
## 12 12 5.22 9.73 4.57 9.03

#7.

ggplot() +  
 geom\_point(aes(x=mean.physical,y=mean.mental,color=Month),  
 data=oh\_summary,alpha=0.5)+  
 labs(x="Monthly Average of Bad Physical Days",  
 y="Monthly Average of Bad Mental Days",  
 title="The Scatter Plot:Average Physical and Mental Days")+  
 theme(plot.title = element\_text(size = 10))

