hw2\_code

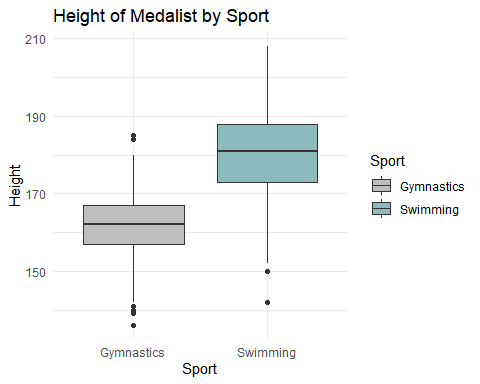
pacman::p\_load(ggplot2, tidyverse, dplyr, plotly, processx)  
setwd('C:/Users/dnskd/Desktop/20Spring/graphics/week2')  
Olympic <- read\_csv('./olympic\_data/athlete\_events.csv')

## Parsed with column specification:  
## cols(  
## ID = col\_double(),  
## Name = col\_character(),  
## Sex = col\_character(),  
## Age = col\_double(),  
## Height = col\_double(),  
## Weight = col\_double(),  
## Team = col\_character(),  
## NOC = col\_character(),  
## Games = col\_character(),  
## Year = col\_double(),  
## Season = col\_character(),  
## City = col\_character(),  
## Sport = col\_character(),  
## Event = col\_character(),  
## Medal = col\_character()  
## )

# Swimming top 10  
sw.noc <- Olympic %>% filter(Sport == "Swimming", !is.na(Medal)) %>% group\_by(NOC, Medal, Year) %>% summarise(n = length(unique(Event))) %>% group\_by(NOC) %>% summarise(total = sum(n)) %>% mutate(n.rank = rank(-total, ties.method = "first")) %>%  
 filter(n.rank < 11)  
  
# Gymnastics top 10  
gm.noc <- Olympic %>% filter(Sport == "Gymnastics", !is.na(Medal)) %>% group\_by(NOC, Medal, Year) %>% summarise(n = length(unique(Event))) %>% group\_by(NOC) %>% summarise(total = sum(n)) %>% mutate(n.rank = rank(-total, ties.method = "first")) %>%  
 filter(n.rank < 11)

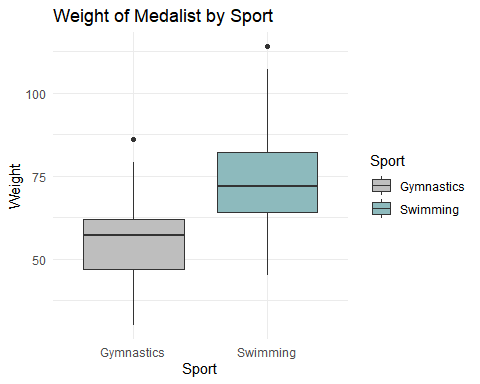
Olympic %>% filter(Sport %in% c("Gymnastics", "Swimming"), !is.na(Medal)) %>%   
 ggplot(aes(Sport, Height, fill = Sport)) + geom\_boxplot() + labs(title = "Height of Medalist by Sport") + theme(plot.title = element\_text(hjust = 0.5, size = 20)) +  
 theme\_bw() + theme\_minimal() + scale\_fill\_manual("Sport", values = c("gray", "#8DBABD"))

## Warning: Removed 1529 rows containing non-finite values (stat\_boxplot).

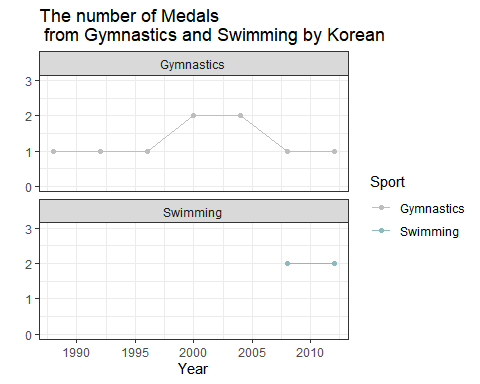


Olympic %>% filter(Sport %in% c("Gymnastics", "Swimming"), !is.na(Medal)) %>%   
 ggplot(aes(Sport, Weight, fill = Sport)) + geom\_boxplot() + labs(title = "Weight of Medalist by Sport") + theme(plot.title = element\_text(hjust = 0.5, size = 20)) +  
 theme\_bw() + theme\_minimal() + scale\_fill\_manual("Sport", values = c("gray", "#8DBABD"))

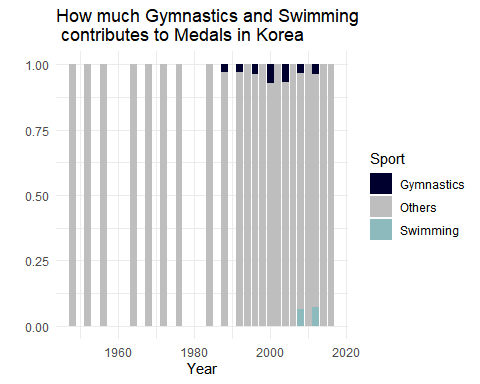
## Warning: Removed 1652 rows containing non-finite values (stat\_boxplot).



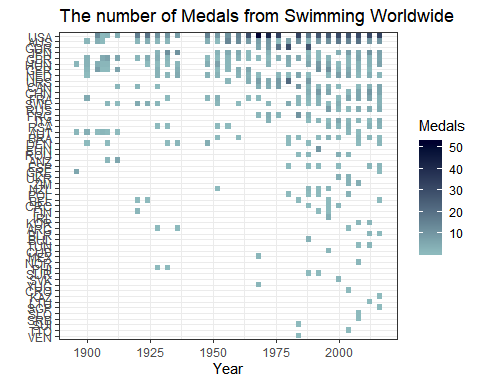
# 그럼 우리나라에 포커스를 두고 Gymnastics와 Swimming 두 개를 비교하자  
Olympic %>% filter(Sport %in% c("Swimming", "Gymnastics"), !is.na(Medal), NOC == "KOR") %>% group\_by(Sport, Year, Medal) %>% summarise(n = length(unique(Event))) %>% group\_by(Sport, Year)%>% summarise(total = sum(n)) %>%  
 ggplot(aes(Year, total, col = Sport)) + geom\_path() + geom\_point() + facet\_wrap(~Sport, ncol = 1) + theme\_bw() + ylim(c(0, 3)) + scale\_color\_manual("Sport", values = c("gray", "#8DBABD"))+ labs(y = " ", title = "The number of Medals \n from Gymnastics and Swimming by Korean")



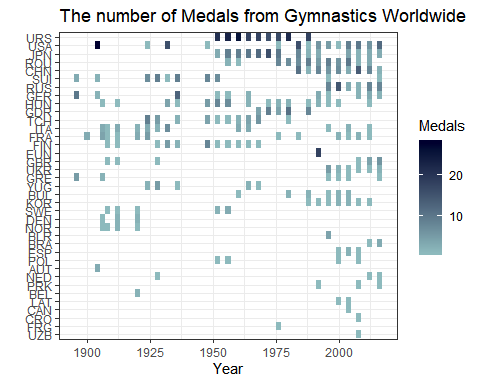
# 전체 메달 중에서 Gymnastic와 Swimming 메달의 개수가 차지하는 비율?  
temp <- Olympic  
temp[which(!temp$Sport %in% c("Swimming", "Gymnastics")), "Sport"] <- "Others"  
temp %>% filter(NOC == "KOR", !is.na(Medal)) %>% group\_by(Sport, Year, Medal) %>% summarise(n = length(unique(Event))) %>% group\_by(Sport, Year) %>% summarise(total = sum(n)) %>%  
 ggplot(aes(Year, total, fill = Sport)) + geom\_bar(stat="identity", position = "fill") + scale\_fill\_manual(values = c("#00002F", "gray", "#8DBABD")) + theme\_bw() + theme\_minimal()+  
 labs(title = "How much Gymnastics and Swimming \n contributes to Medals in Korea", y = " ")



sw.noc <- Olympic %>% filter(Sport == "Swimming", !is.na(Medal)) %>% group\_by(NOC, Medal, Year) %>% summarise(n = length(unique(Event))) %>% group\_by(NOC) %>% summarise(total = sum(n)) %>% mutate(n.rank = rank(-total, ties.method = "first"))  
  
Olympic %>% filter(Sport == "Swimming", !is.na(Medal)) %>% group\_by(NOC, Year, Medal) %>% summarise(n = length(unique(Event))) %>% group\_by(NOC, Year) %>% summarise(total = sum(n)) %>% left\_join(sw.noc, by="NOC") %>% rename(Medals = total.x) %>%  
 ggplot(aes(Year, reorder(NOC, -n.rank), fill = Medals)) + geom\_tile() + labs(title = "The number of Medals from Swimming Worldwide", y = " ") + scale\_fill\_gradient(low = "#8DBABD", high = "#00002F") + theme\_bw()



# Gymnastics  
gm.noc <- Olympic %>% filter(Sport == "Gymnastics", !is.na(Medal)) %>% group\_by(NOC, Medal, Year) %>% summarise(n = length(unique(Event))) %>% group\_by(NOC) %>% summarise(total = sum(n)) %>% mutate(n.rank = rank(-total, ties.method = "first"))  
  
Olympic %>% filter(Sport == "Gymnastics", !is.na(Medal)) %>% group\_by(NOC, Year, Medal) %>% summarise(n = length(unique(Event))) %>% group\_by(NOC, Year) %>% summarise(total = sum(n)) %>% left\_join(gm.noc, by="NOC") %>% rename(Medals = total.x) %>%  
 ggplot(aes(Year, reorder(NOC, -n.rank), fill = Medals)) + geom\_tile() + labs(title = "The number of Medals from Gymnastics Worldwide", y = " ") + scale\_fill\_gradient(low = "#8DBABD", high = "#00002F") + theme\_bw()



sw.top <- sw.noc %>% filter(n.rank < 7) %>% dplyr::select("NOC") %>% unlist()  
gm.top <- gm.noc %>% filter(n.rank < 7) %>% dplyr::select("NOC") %>% unlist()  
# Swimming  
library(vcd); library(HH)

## Warning: package 'vcd' was built under R version 3.6.1

## Loading required package: grid

## Warning: package 'HH' was built under R version 3.6.3

## Loading required package: lattice

## Warning: package 'lattice' was built under R version 3.6.1

## Loading required package: latticeExtra

## Warning: package 'latticeExtra' was built under R version 3.6.1

## Loading required package: RColorBrewer

##   
## Attaching package: 'latticeExtra'

## The following object is masked from 'package:vcd':  
##   
## rootogram

## The following object is masked from 'package:ggplot2':  
##   
## layer

## Loading required package: multcomp

## Warning: package 'multcomp' was built under R version 3.6.3

## Loading required package: mvtnorm

## Loading required package: survival

## Loading required package: TH.data

## Warning: package 'TH.data' was built under R version 3.6.3

## Loading required package: MASS

##   
## Attaching package: 'MASS'

## The following object is masked from 'package:plotly':  
##   
## select

## The following object is masked from 'package:dplyr':  
##   
## select

##   
## Attaching package: 'TH.data'

## The following object is masked from 'package:MASS':  
##   
## geyser

## Loading required package: gridExtra

##   
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':  
##   
## combine

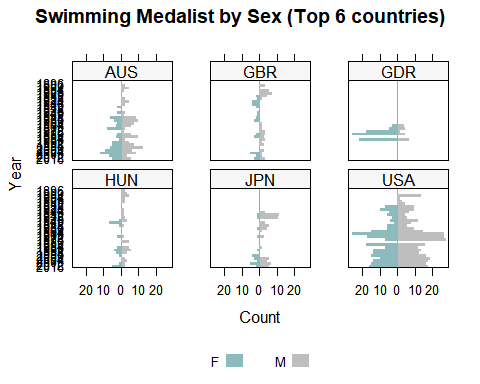
## This version of Shiny is designed to work with 'htmlwidgets' >= 1.5.  
## Please upgrade via install.packages('htmlwidgets').

##   
## Attaching package: 'HH'

## The following object is masked from 'package:vcd':  
##   
## odds

## The following object is masked from 'package:purrr':  
##   
## transpose

sw1 <- Olympic %>% filter(Sport == "Swimming", !is.na(Medal), NOC %in% sw.top) %>% group\_by(NOC, Sex, Medal, Year) %>% summarise(n = length(unique(Event))) %>% group\_by(NOC, Sex, Year) %>% summarise(total = sum(n)) %>%  
 spread(Sex, value = total) %>% as.data.frame()  
  
insertftn <- function(origindf, r, year){  
 origindf[seq((r+1),(nrow(origindf)+1)), ] <- origindf[seq(r,nrow(origindf)), ]  
 origindf[r, ] <- data.frame(NOC = "AUS", Year = year, F = NA, M = NA)  
 return(origindf)  
}  
  
# 1896  
a <- insertftn(sw1, r = 1, year = 1896 )  
# 1908  
b <- insertftn(a, r = 5, year = 1908)  
# 1912  
c <- insertftn(b, r = 6, year = 1912)  
# 1936  
final <- insertftn(c, r = 11, year = 1936)  
  
final[final$NOC==1, 'NOC'] <- 'AUS'  
  
likert(Year ~ . | NOC, final, main = "Swimming Medalist by Sex (Top 6 countries)", col = c("#8DBABD", "gray"))



# Gymnastics  
gm1 <- Olympic %>% filter(Sport == "Gymnastics", !is.na(Medal), NOC %in% gm.top) %>% group\_by(NOC, Sex, Medal, Year) %>% summarise(n = length(unique(Event))) %>% group\_by(NOC, Sex, Year) %>% summarise(total = sum(n)) %>%  
 spread(Sex, value = total) %>% as.data.frame()  
subset <- gm1[seq(55, nrow(gm1)), ]  
subset2 <- gm1[seq(1, 54), ]  
gm2 <- rbind(subset, subset2)  
rownames(gm2) <- seq(1:nrow(gm2))  
  
# 1896  
a <- insertftn(gm2, r = 1, year = 1896 )  
head(a)

## NOC Year F M  
## 1 1 1896 NA NA  
## 2 USA 1904 NA 28  
## 3 USA 1924 NA 1  
## 4 USA 1932 NA 16  
## 5 USA 1948 1 NA  
## 6 USA 1976 NA 1

# 1928  
b <- insertftn(a, r = 4, year = 1928 )  
head(b)

## NOC Year F M  
## 1 1 1896 NA NA  
## 2 USA 1904 NA 28  
## 3 USA 1924 NA 1  
## 4 1 1928 NA NA  
## 5 USA 1932 NA 16  
## 6 USA 1948 1 NA

# 1936  
c <- insertftn(b, r = 6, year = 1936 )  
head(c, n = 10)

## NOC Year F M  
## 1 1 1896 NA NA  
## 2 USA 1904 NA 28  
## 3 USA 1924 NA 1  
## 4 1 1928 NA NA  
## 5 USA 1932 NA 16  
## 6 1 1936 NA NA  
## 7 USA 1948 1 NA  
## 8 USA 1976 NA 1  
## 9 USA 1984 8 8  
## 10 USA 1988 1 NA

# 1952  
d <- insertftn(c, r = 8, year = 1952 )  
head(d, n = 15)

## NOC Year F M  
## 1 1 1896 NA NA  
## 2 USA 1904 NA 28  
## 3 USA 1924 NA 1  
## 4 1 1928 NA NA  
## 5 USA 1932 NA 16  
## 6 1 1936 NA NA  
## 7 USA 1948 1 NA  
## 8 1 1952 NA NA  
## 9 USA 1976 NA 1  
## 10 USA 1984 8 8  
## 11 USA 1988 1 NA  
## 12 USA 1992 5 1  
## 13 USA 1996 4 1  
## 14 USA 2000 1 NA  
## 15 USA 2004 6 3

# 1956  
e <- insertftn(d, r = 9, year = 1956 )  
head(e, n = 15)

## NOC Year F M  
## 1 1 1896 NA NA  
## 2 USA 1904 NA 28  
## 3 USA 1924 NA 1  
## 4 1 1928 NA NA  
## 5 USA 1932 NA 16  
## 6 1 1936 NA NA  
## 7 USA 1948 1 NA  
## 8 1 1952 NA NA  
## 9 1 1956 NA NA  
## 10 USA 1976 NA 1  
## 11 USA 1984 8 8  
## 12 USA 1988 1 NA  
## 13 USA 1992 5 1  
## 14 USA 1996 4 1  
## 15 USA 2000 1 NA

# 1960  
f <- insertftn(e, r = 10, year = 1960 )  
head(f, n = 15)

## NOC Year F M  
## 1 1 1896 NA NA  
## 2 USA 1904 NA 28  
## 3 USA 1924 NA 1  
## 4 1 1928 NA NA  
## 5 USA 1932 NA 16  
## 6 1 1936 NA NA  
## 7 USA 1948 1 NA  
## 8 1 1952 NA NA  
## 9 1 1956 NA NA  
## 10 1 1960 NA NA  
## 11 USA 1976 NA 1  
## 12 USA 1984 8 8  
## 13 USA 1988 1 NA  
## 14 USA 1992 5 1  
## 15 USA 1996 4 1

# 1964  
g <- insertftn(f, r = 11, year = 1964 )  
head(g, n = 15)

## NOC Year F M  
## 1 1 1896 NA NA  
## 2 USA 1904 NA 28  
## 3 USA 1924 NA 1  
## 4 1 1928 NA NA  
## 5 USA 1932 NA 16  
## 6 1 1936 NA NA  
## 7 USA 1948 1 NA  
## 8 1 1952 NA NA  
## 9 1 1956 NA NA  
## 10 1 1960 NA NA  
## 11 1 1964 NA NA  
## 12 USA 1976 NA 1  
## 13 USA 1984 8 8  
## 14 USA 1988 1 NA  
## 15 USA 1992 5 1

# 1968  
h <- insertftn(g, r = 12, year = 1968 )  
head(h, n = 15)

## NOC Year F M  
## 1 1 1896 NA NA  
## 2 USA 1904 NA 28  
## 3 USA 1924 NA 1  
## 4 1 1928 NA NA  
## 5 USA 1932 NA 16  
## 6 1 1936 NA NA  
## 7 USA 1948 1 NA  
## 8 1 1952 NA NA  
## 9 1 1956 NA NA  
## 10 1 1960 NA NA  
## 11 1 1964 NA NA  
## 12 1 1968 NA NA  
## 13 USA 1976 NA 1  
## 14 USA 1984 8 8  
## 15 USA 1988 1 NA

# 1972  
new1 <- insertftn(h, r = 13, year = 1972 )  
head(new1, n = 15)

## NOC Year F M  
## 1 1 1896 NA NA  
## 2 USA 1904 NA 28  
## 3 USA 1924 NA 1  
## 4 1 1928 NA NA  
## 5 USA 1932 NA 16  
## 6 1 1936 NA NA  
## 7 USA 1948 1 NA  
## 8 1 1952 NA NA  
## 9 1 1956 NA NA  
## 10 1 1960 NA NA  
## 11 1 1964 NA NA  
## 12 1 1968 NA NA  
## 13 1 1972 NA NA  
## 14 USA 1976 NA 1  
## 15 USA 1984 8 8

# 1980  
new2 <- insertftn(new1, r = 15, year = 1980 )  
head(new2, n = 15)

## NOC Year F M  
## 1 1 1896 NA NA  
## 2 USA 1904 NA 28  
## 3 USA 1924 NA 1  
## 4 1 1928 NA NA  
## 5 USA 1932 NA 16  
## 6 1 1936 NA NA  
## 7 USA 1948 1 NA  
## 8 1 1952 NA NA  
## 9 1 1956 NA NA  
## 10 1 1960 NA NA  
## 11 1 1964 NA NA  
## 12 1 1968 NA NA  
## 13 1 1972 NA NA  
## 14 USA 1976 NA 1  
## 15 1 1980 NA NA

new2[which(new2$NOC == 1), "NOC"] <- 'USA'  
likert(Year ~ . | NOC, new2, main = "Gymnastics Medalist by Sex (Top 6 countries)", col = c("#8DBABD", "gray"))

