

tut10

R Markdown

Heights seem to be important in certain sports, i.e. Basketball players are taller and Gymnasts are shorter. Did all countries realized the importance of heights in these sports at the beginning of Olympics,or was it only after a certain period?

First,we create a line plot comparing the mean heights of all the athletes across the years.Maybe we should split them by their gender too as that might affect the mean height of the athletes.(Gender and Height may have an association)

```
library(tidyverse)

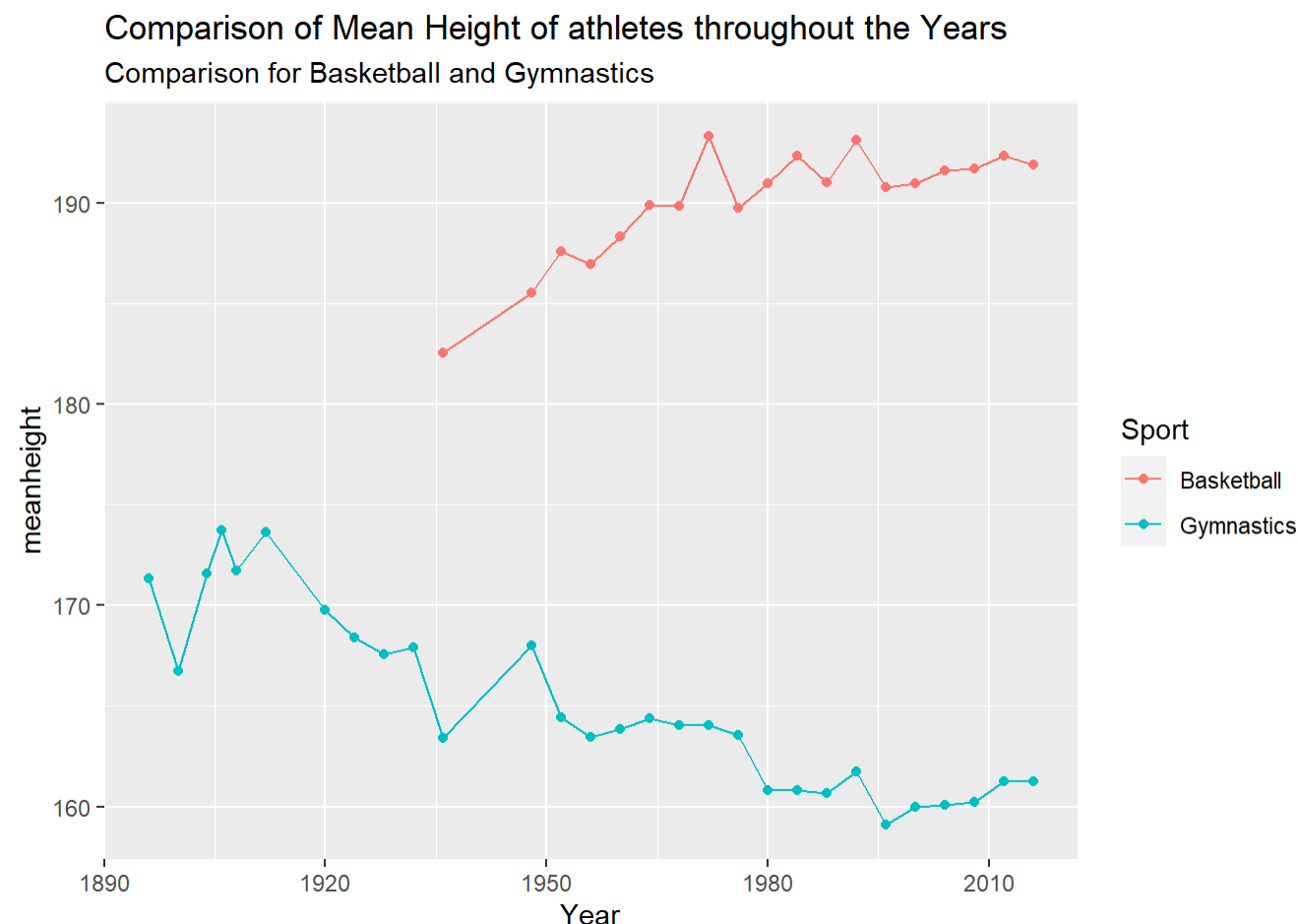
## -- Attaching packages ----- tidyverse 1.3.0 --

## v ggplot2 3.3.3    v purrr  0.3.4
## v tibble  3.0.5    v dplyr   1.0.3
## v tidyr   1.1.2    v stringr 1.4.0
## v readr   1.4.0    v forcats 0.5.1

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

olympics<-read.csv("../data/athlete_events.csv")
tmp<-olympics%>%
  filter(!is.na(Height))%>%
  filter(Sport %in% c("Basketball","Gymnastics"))
tmp%>%group_by(Year)%>%
  distinct_at(vars(ID),.keep_all=TRUE)%>%
  ungroup()%>%
  group_by(Year,Sport)%>%
  summarize(meanheight = mean(Height))%>%
ggplot(aes(x=Year,y=meanheight,group=Sport,color=Sport))+
  geom_point()+
  geom_line()+
  labs(title="Comparison of Mean Height of athletes throughout the Years",subtitle = "Comparison for Basketball and Gymnastics")

## `summarise()` has grouped output by 'Year'. You can override using the `.groups` argument.
```

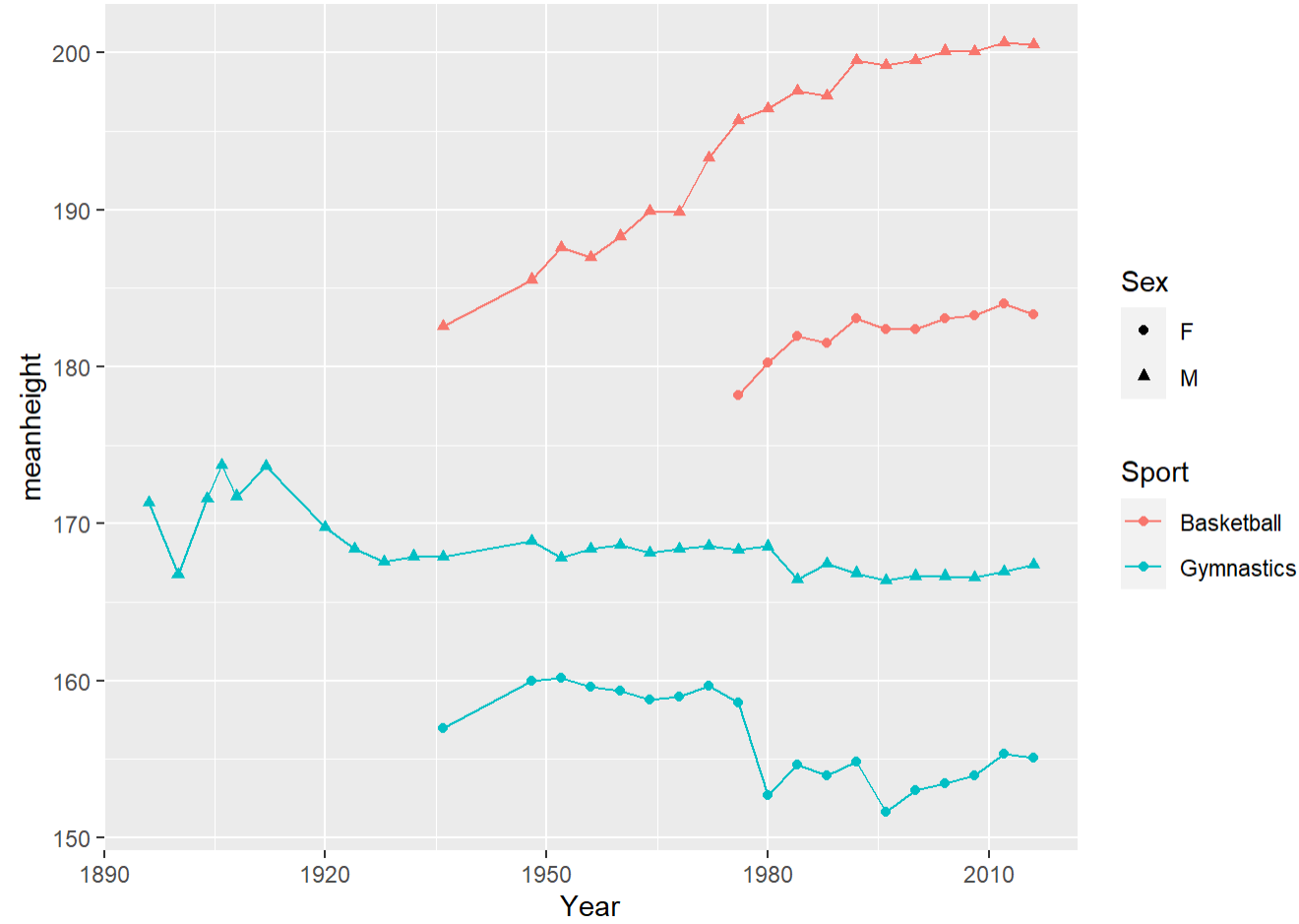


```
#data<-tmp%>%group_by(Year)%>%
#distinct(tmp$ID)
```

There might be an association between sex and height so we plot the following below:

```
tmp%>%group_by(Year)%>%
  distinct_at(vars(ID),.keep_all=TRUE)%>%
  ungroup()%>%
  group_by(Sex,Year,Sport)%>%
  summarize(meanheight=mean(Height))%>%
ggplot(aes(x=Year,y=meanheight,group=interaction(Sport,Sex),color=Sport,shape = Sex))+
  geom_point()+
  geom_line()

## `summarise()` has grouped output by 'Sex', 'Year'. You can override using the `.groups` argument.
```



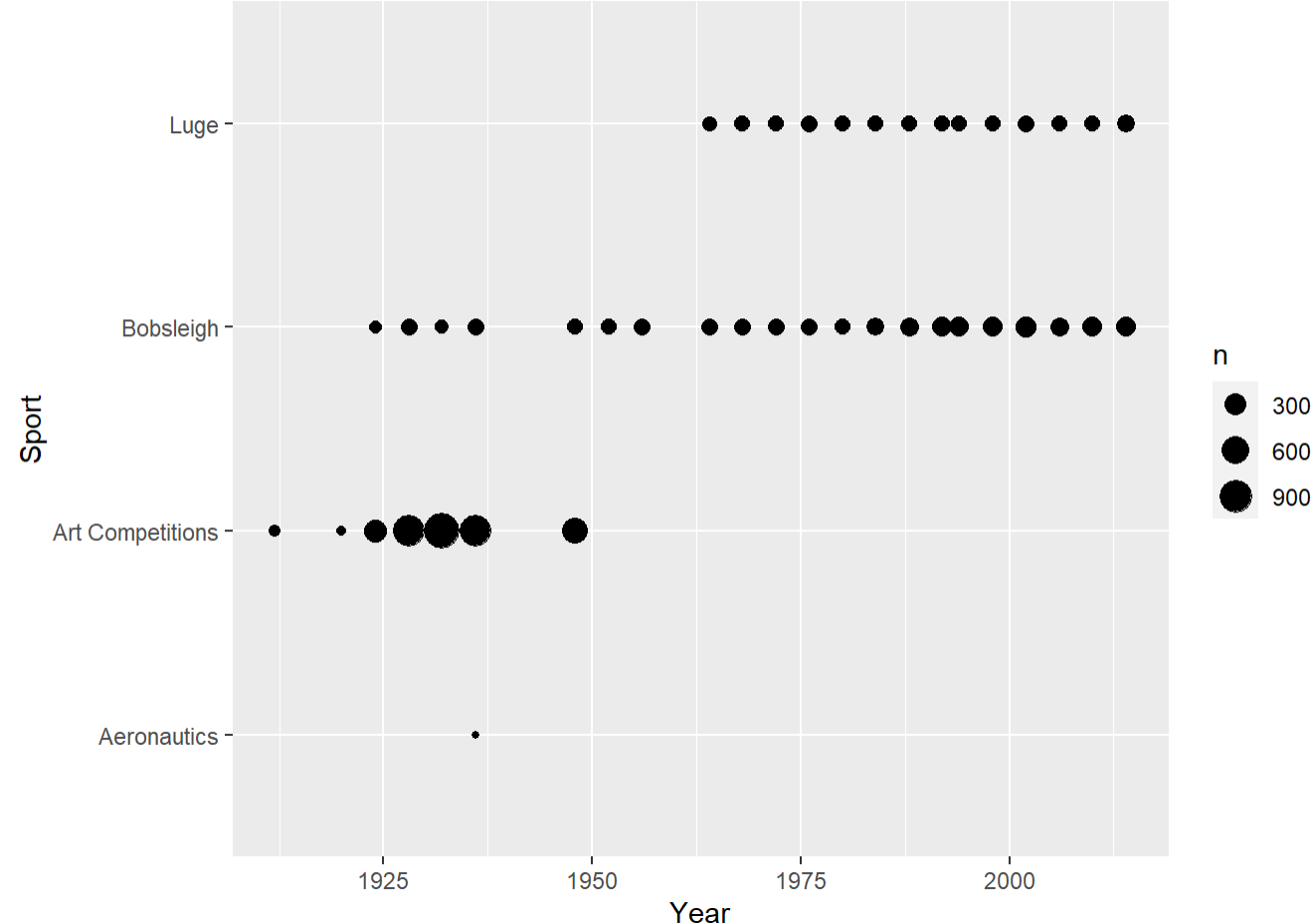
From the plots above,gymnastics and Basketball for females started at a later time as compared to Male Teams for Basketball and Gymnastics. By observing the trends above, the heights of Basketball players is increasing over the years for both males and females. However, for gymnastics we see an increase in heights between 1900 and 1908 and then on starts to decline once again.(For Males) For Female gymnastic athletes, we see a general decreasing trend in height. However, there is a very sharpest dip is seen in 1980. We can go on to inspect this to see if this sharpest dip is because of only few gymnast athletes participating in that year. But that does not seem to be the case from our code below.

```
olympics%>%
  filter(!is.na(Height))%>%
  filter(Year %in% c(1980))%>%
  filter(Sport=="Gymnastics")%>%
  group_by(Year)%>%
  distinct_at(vars(ID),.keep_all=TRUE)

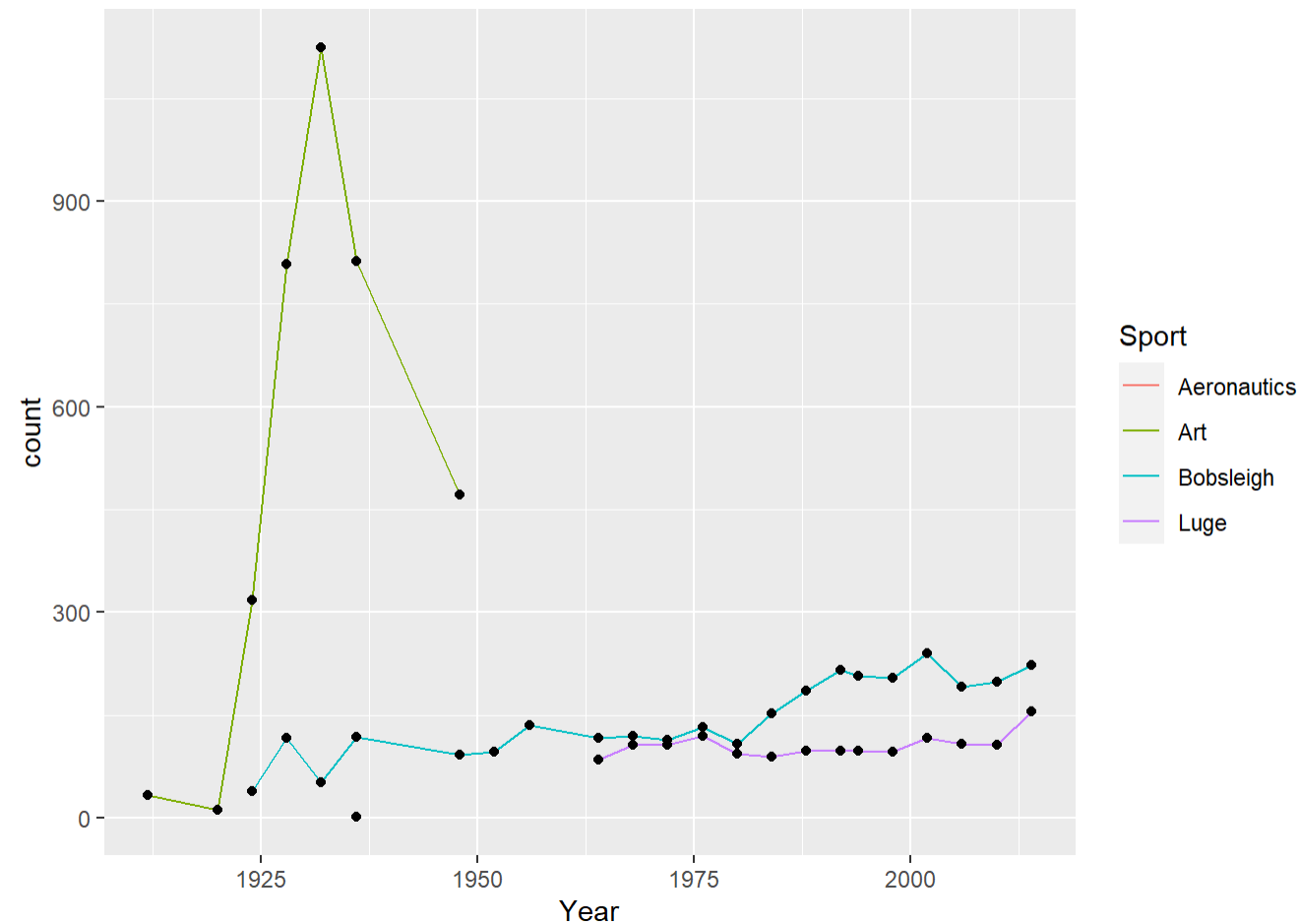
## # A tibble: 127 x 15
## # Groups:   Year [1]
##   ID Name Sex Age Height Weight Team NOC Games Year Season City
##   <int> <chr> <chr> <int> <int> <dbl> <chr> <chr> <chr> <int> <chr> <chr>
## 1 1003 Lena~ F 16 162 56 Swed~ SWE 1980~ 1980 Summer Mosk~
## 2 2959 Lenk~ F 15 148 40 Hung~ HUN 1980~ 1980 Summer Mosk~
## 3 3142 Mari~ F 18 162 45 Port~ POR 1980~ 1980 Summer Mosk~
## 4 4198 Niko~ M 27 166 60 Sovi~ URS 1980~ 1980 Summer Mosk~
## 5 5322 Migu~ M 18 165 54 Cuba CUB 1980~ 1980 Summer Mosk~
## 6 6301 Edua~ M 22 175 68 Sovi~ URS 1980~ 1980 Summer Mosk~
## 7 6449 Rudo~ M 24 157 53 Czec~ TCH 1980~ 1980 Summer Mosk~
## 8 7509 Ogn~ M 24 169 63 Bulg~ BUL 1980~ 1980 Summer Mosk~
## 9 8914 Kerr~ F 18 157 41 Aust~ AUS 1980~ 1980 Summer Mosk~
## 10 10918 Fern~ M 25 170 67 Spain ESP 1980~ 1980 Summer Mosk~
## # ... with 117 more rows, and 3 more variables: Sport <chr>, Event <chr>,
## # Medal <chr>
```

There are some interesting sports events over the years of the Olympics, such as Art Competitions,Luge, Bobsleigh and Aeronautics. However, some of these are no longer Olympic sports. In general,how has the popularity of the sports events changed over the years?

```
olympics%>%
  filter(Sport %in% c("Art Competitions","Luge","Bobsleigh","Aeronautics"))%>%
  group_by(Year,Sport)%>%
  ggplot()+
  geom_count(aes(x=Year,y=Sport))
```



```
olympics%>%
  filter(Sport %in% c("Art Competitions","Luge","Bobsleigh","Aeronautics"))%>%
  group_by(Year,Sport)%>%
  mutate(count=n())%>%
  mutate(Sport=recode(Sport,"Art Competitions"="Art"))%>%
  ggplot()+
  geom_line(aes(x=Year,y=count,group=Sport,color=Sport))+
  geom_point(aes(x=Year,y=count))
```



Since size may be visually hard to compare the counts , I modify my graph into a line plot. The popularity of Arts Competition rose from 1924 to 1932 and then started dropping once again. Aeronautics only occurred for one year. The Bobsleigh seemed to be having a very slight increase over the years while the Luge seemed to be quite stagnant over the years with small decreaseas and increases in the numbers