

Übung 05

Grafische Darstellung mit R

Erste Anwendung von **qplot()**

INFI-IS

5AHWII

Albert Greinöcker, Thomas Kefer

December 12, 2021



1 Datensatz: Olympische Spiele

Dieser historische Datensatz beinhaltet alle Ergebnisse der Olympischen Spiele von Athen 1896 bis Rio 2016.
Quelle(n): www.kaggle.com bzw. <https://www.sports-reference.com>.

1.1 Importieren des Datensatzes

Die entsprechende zip-Datei entpacken und wie gehabt importieren:

```
1 setwd('/home/baumgart13/Desktop/Mitschrift_Lokal/Schule/GITHUB-ONLINE/Mitschrift/5/INFI/
  IS/Aufgabe05/')
2 o = read.csv("../datasets/athlete_events.csv", sep=";", header=T)
```

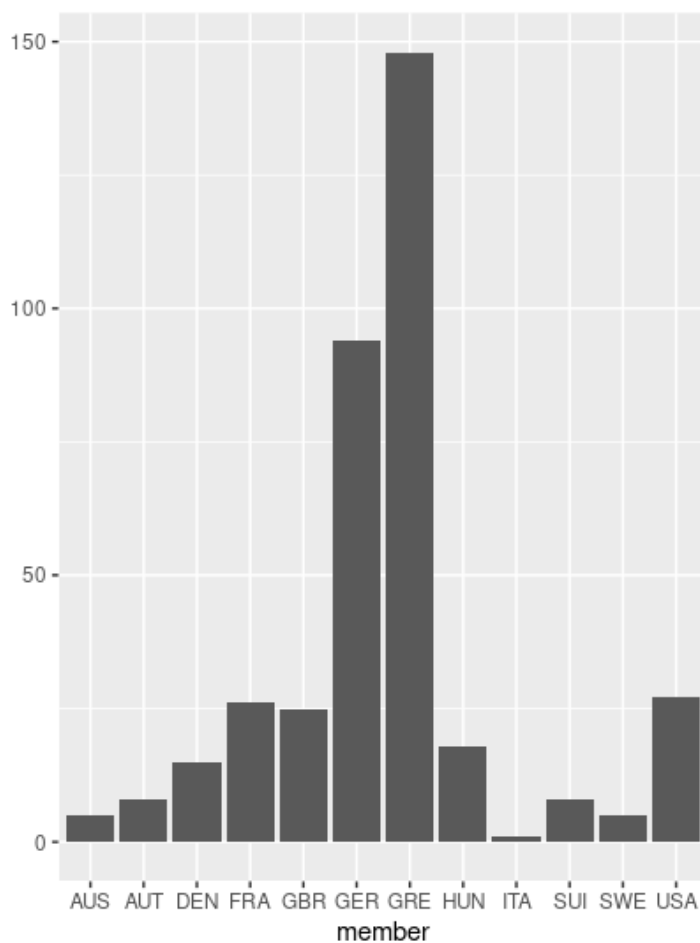
Insgesamt sind es 271116 Einträge. Die einzelnen Variablen sind durch ihre Namensgebung selbsterklärend.

2 Bitte folgende Fragestellungen grafisch darstellen

Für die Grafiken soll `qplot()` aus dem Paket `ggplot2` verwendet werden. Ähnliche Beispiele haben wir mit einem anderen Datensatz (`student`) gemeinsam gemacht. Ein entsprechendes Skript dazu ist in Moodle zu finden.

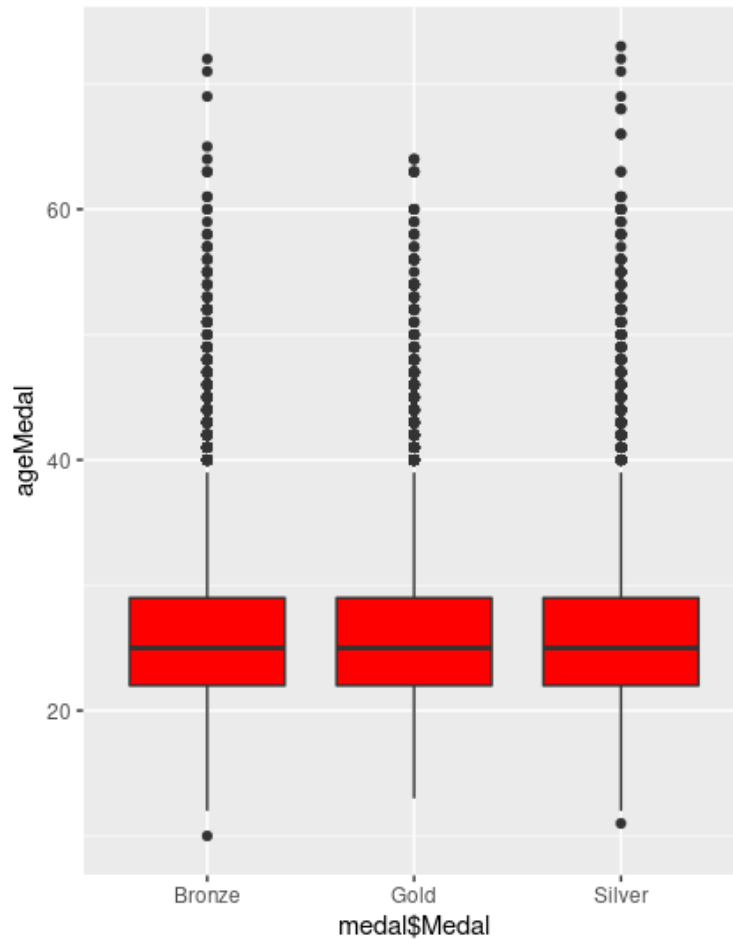
- a. Wie viele Teilnehmer gab es in den einzelnen Ländern im Jahr 1896? Hier wäre ein `barplot` sinnvoll. Am Besten hier das Länderkürzel (NOC) verwenden.

```
1 member <- o$NOC[as.numeric(o$Year) == "1896"]  
2 qplot(member, geom=c("bar"))
```



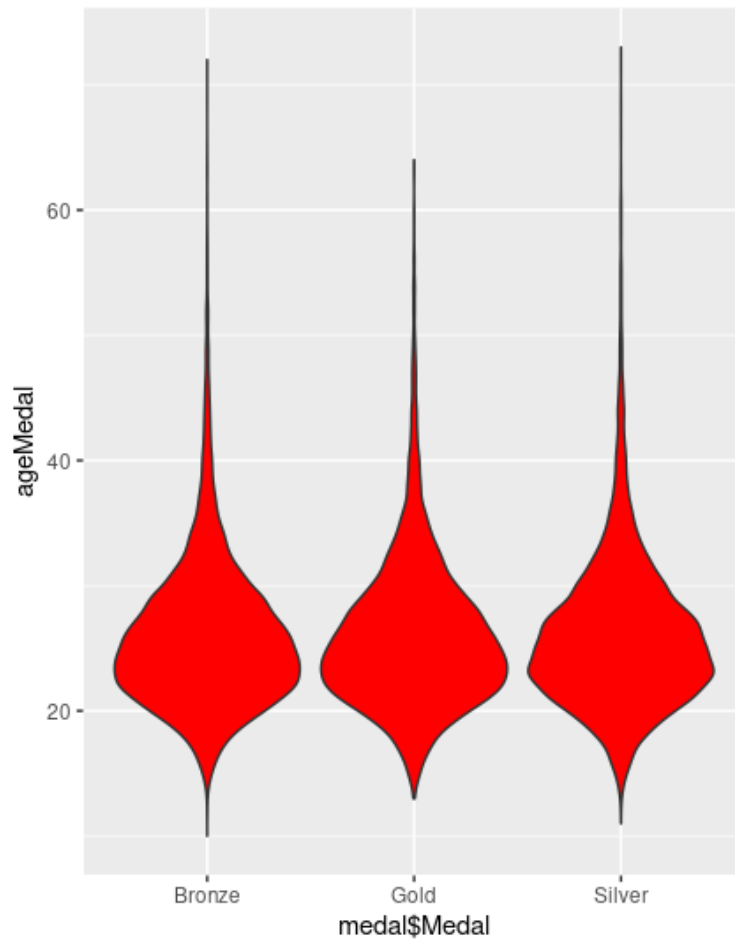
- b. Altersverteilung der einzelnen Medaillengewinner. Am Besten soll ein Boxplot erstellt werden. Als Füllfarbe (fill) soll rot (`I("red")`) gesetzt werden. Die Darstellung soll auf vertikal umgestellt werden: `coord_flip()`.

```
1 ageMedal <- o$Age[(o$Medal) %in% c("Bronze", "Silver", "Gold")]
2 medal <- o[(o$Medal) %in% c("Bronze", "Silver", "Gold"),]
3 qplot(ageMedal, medal$Medal, geom=c("boxplot"), fill=(I("red"))) + coord_flip()
```



- c. Dasselbe bitte als `violin`-plot. Welchen Informationsgewinn bekommen wir bei dieser Darstellungsform?

```
1 qplot(ageMedal, medal$Medal, geom=c("violin"), fill=(I("red"))) + coord_flip()
```



- d. Theoretische Frage (muss nicht in R gemacht werden, sondern der Lösungsansatz nur beschrieben): Die Ordnung der einzelnen Grafiken ist jetzt Bronze, Gold, Silber. Wie bekommt man die richtige Ordnung Bronze, Silber, Gold?

Hinweis: Es hat etwas mit dem Datentyp zu tun.

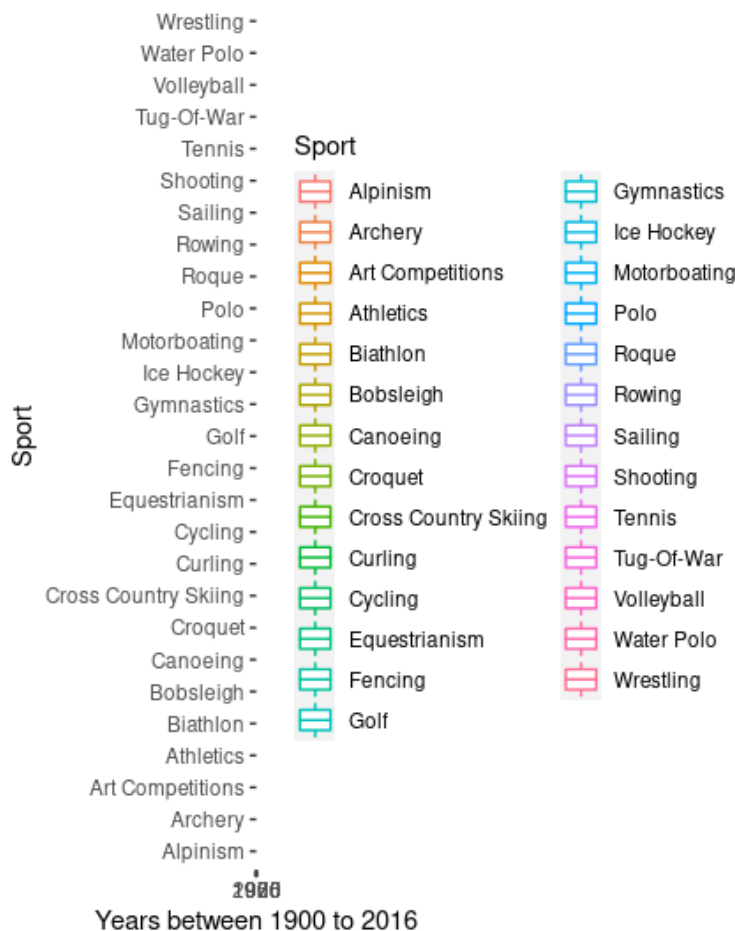
```
1 factor(o$Medal, levels=c("Bronze", "Silver", "Gold"))
```

- e. Es gibt einige, die über 40 sind und eine Goldmedaille gewonnen haben. Interessant wäre, um welche Sportarten es sich handelt.

Es ist also notwendig, eine Auswahl aller über 40-Jährigen zu treffen, gemeinsam mit denen die Medaillen gewonnen haben.

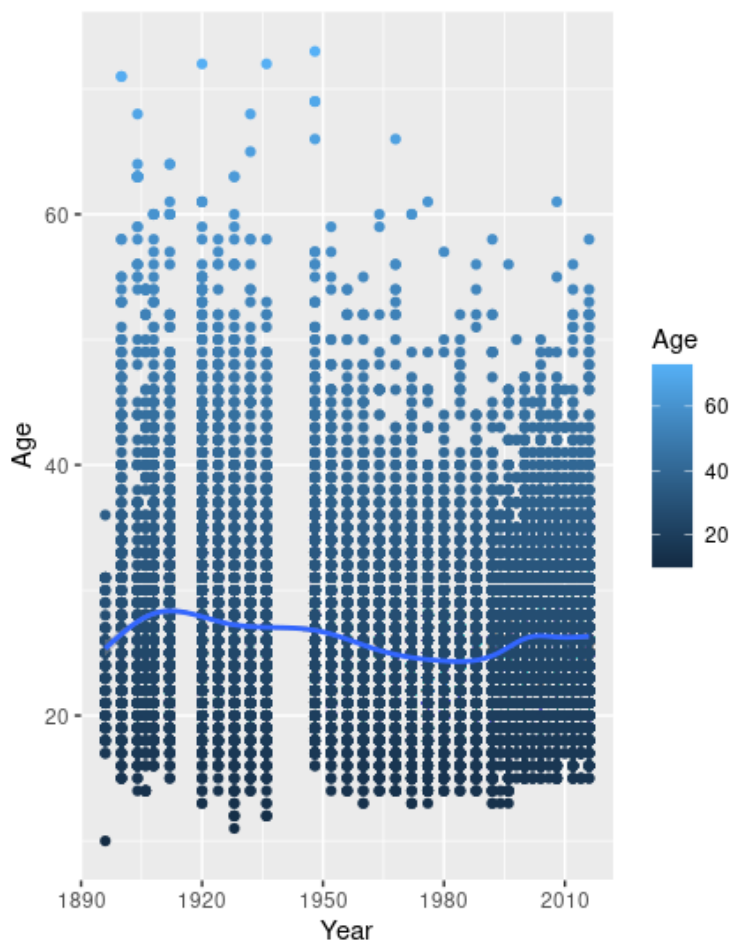
Hinweis: Für die Auswahl derer, die keine Medaille gewonnen haben, kann die Funktion `is.na` verwendet werden, also: `is.na(o$Medal)` Es soll auf der X-Achse (Ticks-Beschriftung alle 2 Jahre) die Jahre Beschriftet werden, auf Y die Sportarten, und die Medaillen in unterschiedlichen Farben. Mögliche Darstellung: Punktdiagramm.

```
1 goldMedal <- o[(o$Medal) %in% "Gold" ,]
2 goldMedal_40_up <- goldMedal[as.numeric(goldMedal$Age) >= 40 & !is.na(goldMedal$
3   Age) ,]
4 goldMedal_40_up_sport <- goldMedal_40_up$Sport
   qplot(Year, Sport, geom=c("boxplot"), xlab="Years between 1900 to 2016", col=Sport ,
       data=goldMedal_40_up)
```



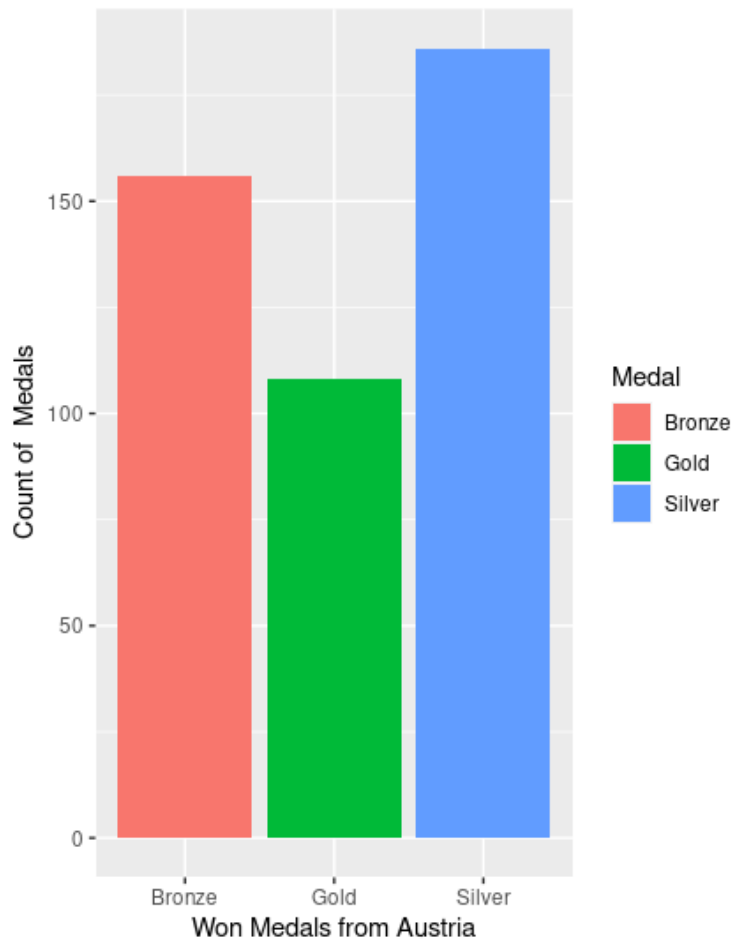
- f. Es wäre auch noch wünschenswert zu sehen ob sich das hohe Alter der Medaillengewinner über die Jahre verändert hat, d.h. ob die Zahl der älteren Gewinner über die Jahre abnimmt.

```
1 medalAge <- medal[!is.na(medal$Age),]  
2 qplot(Year, Age, geom=c("point", "smooth"), col=Age, data=medalAge)
```



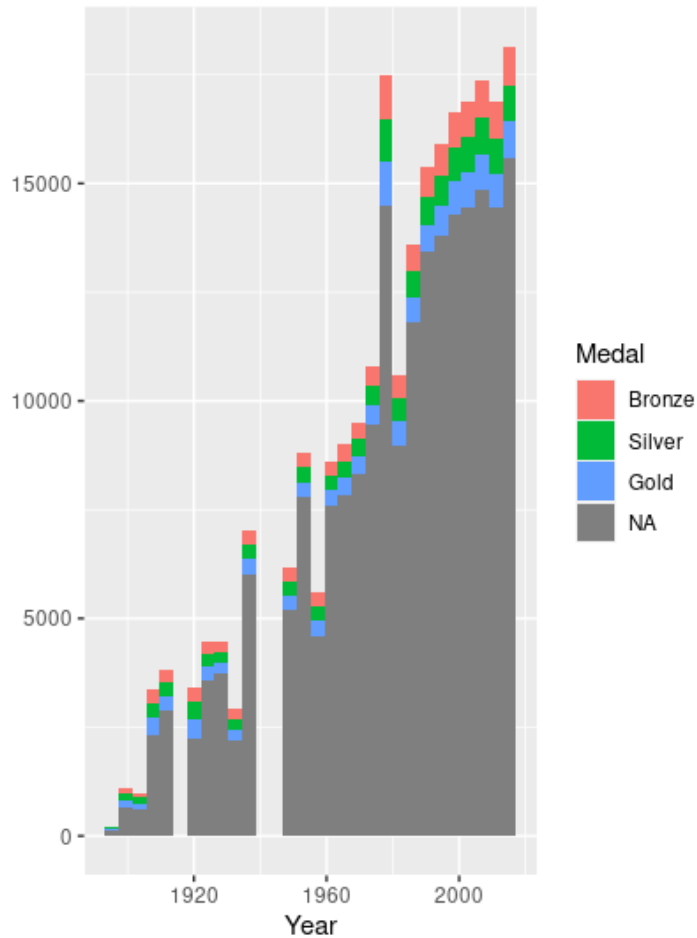
- g. Wenn wir jetzt nur die österreichischen Teilnehmer betrachten, dann wäre die Medaillenentwicklung interessant. Es ist also notwendig, nur die österreichischen Teilnehmer mit Medaillen herauszufiltern und als Balkendiagramm darzustellen. Welche Medaille gewonnen wurde soll farblich im Balken dargestellt werden.

```
1 medalAt <- medal[medal$NOC == "AUT",]  
2 qplot(medalAt$Medal, geom=c("bar"), xlab="Won Medals from Austria", ylab="Count of  
Medals", data = medalAt, fill = Medal)
```



- h. Bitte selbst noch eine Fragestellung finden und das Ergebnis grafisch darstellen. Wie ist die Verteilung der Medaillen unter den Erwachsenen Teilnehmer? Wie viele haben überhaupt keine Medaillen erhalten?

```
1 grownUps <- o[o$Age >= 18,]  
2 qplot(Year, geom=c("histogram"), data=grownUps, fill=Medal)
```



Natürlich sollen alle Grafiken exportiert, ins Dokument eingefügt und interpretiert werden.