Lab 1 Analysis

Stephen Cook

August 27 2018

Analysis

Import packages

```
require(tidyverse)
require(stringr)
```

Import data

```
human <- read.csv(file = "human_data.csv", header = T, sep = ",")
```

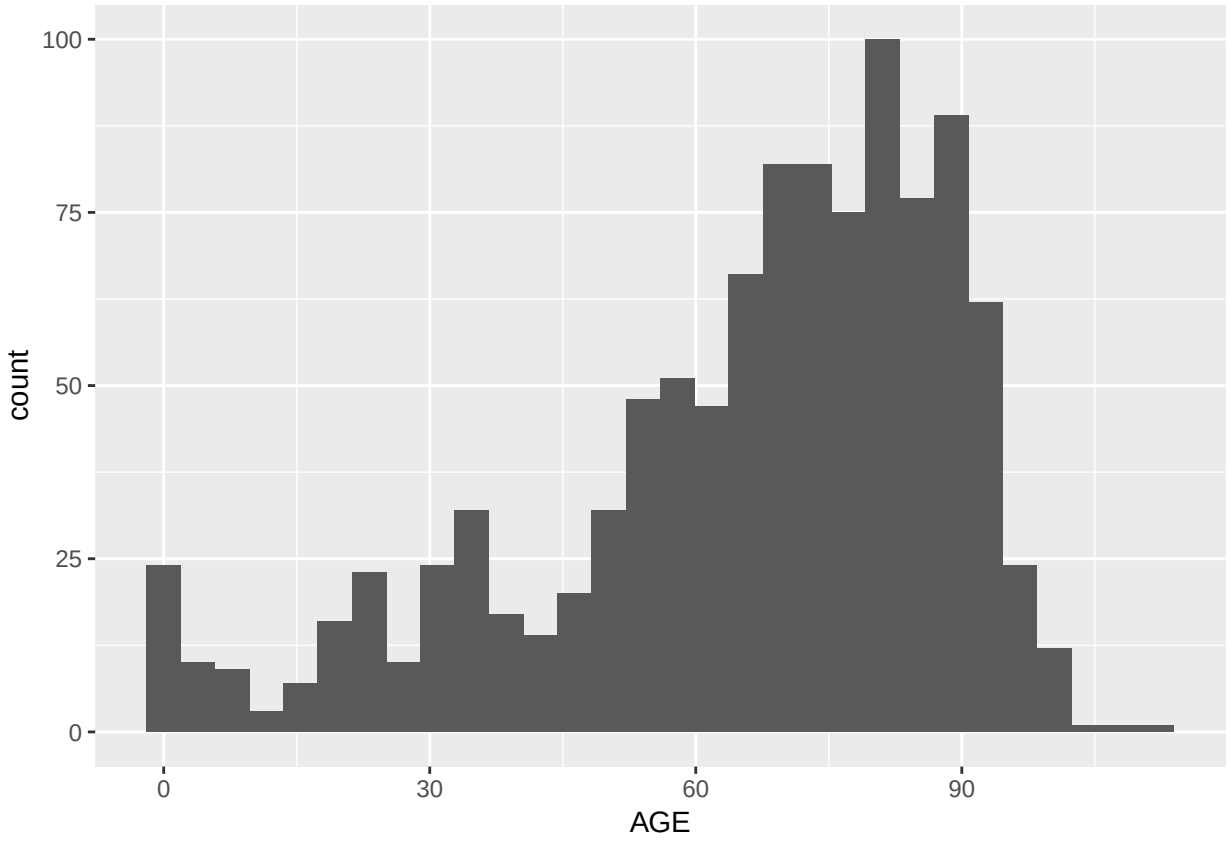
Calculations

```
human <- human %>%
  mutate(AGE = DEATH - BIRTH)
```

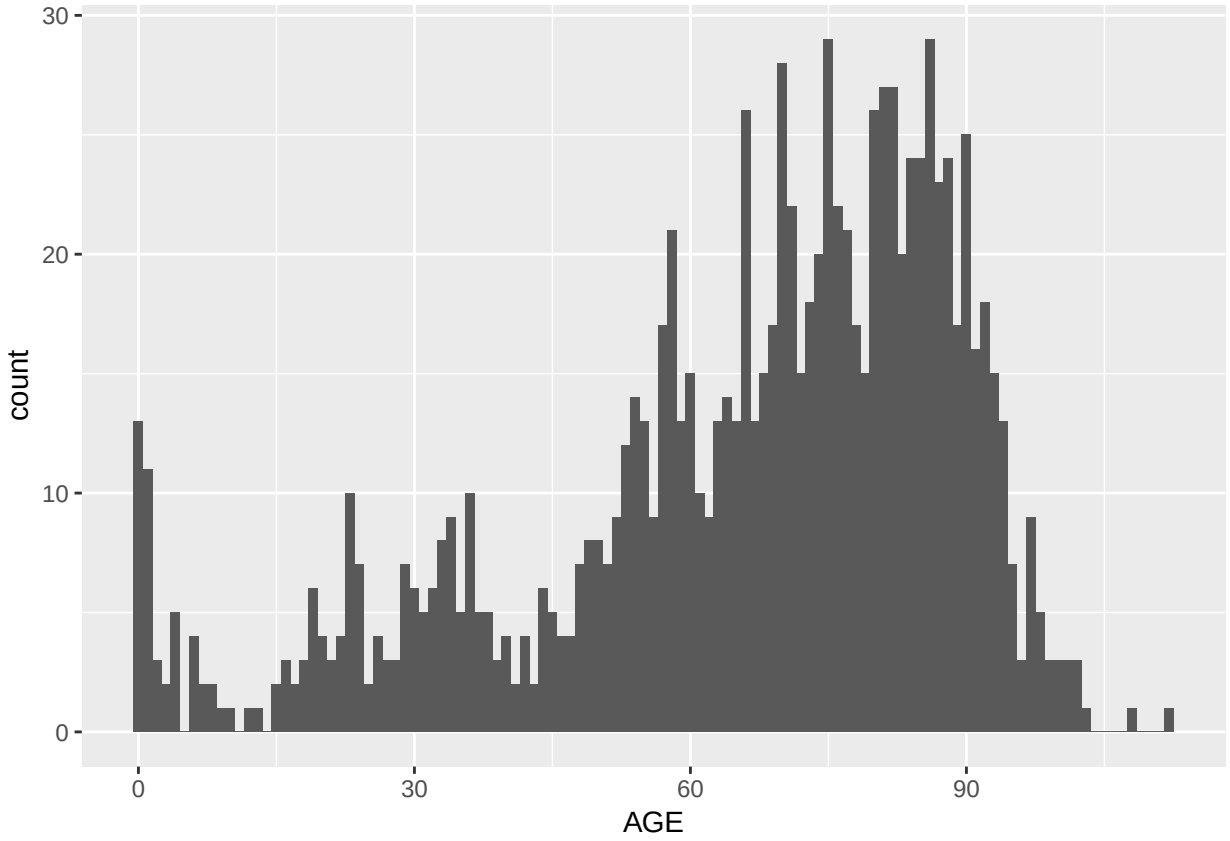
Plotting

```
plot1 <- ggplot() +
  geom_histogram(data = human,
                 aes(x = AGE))

plot1
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



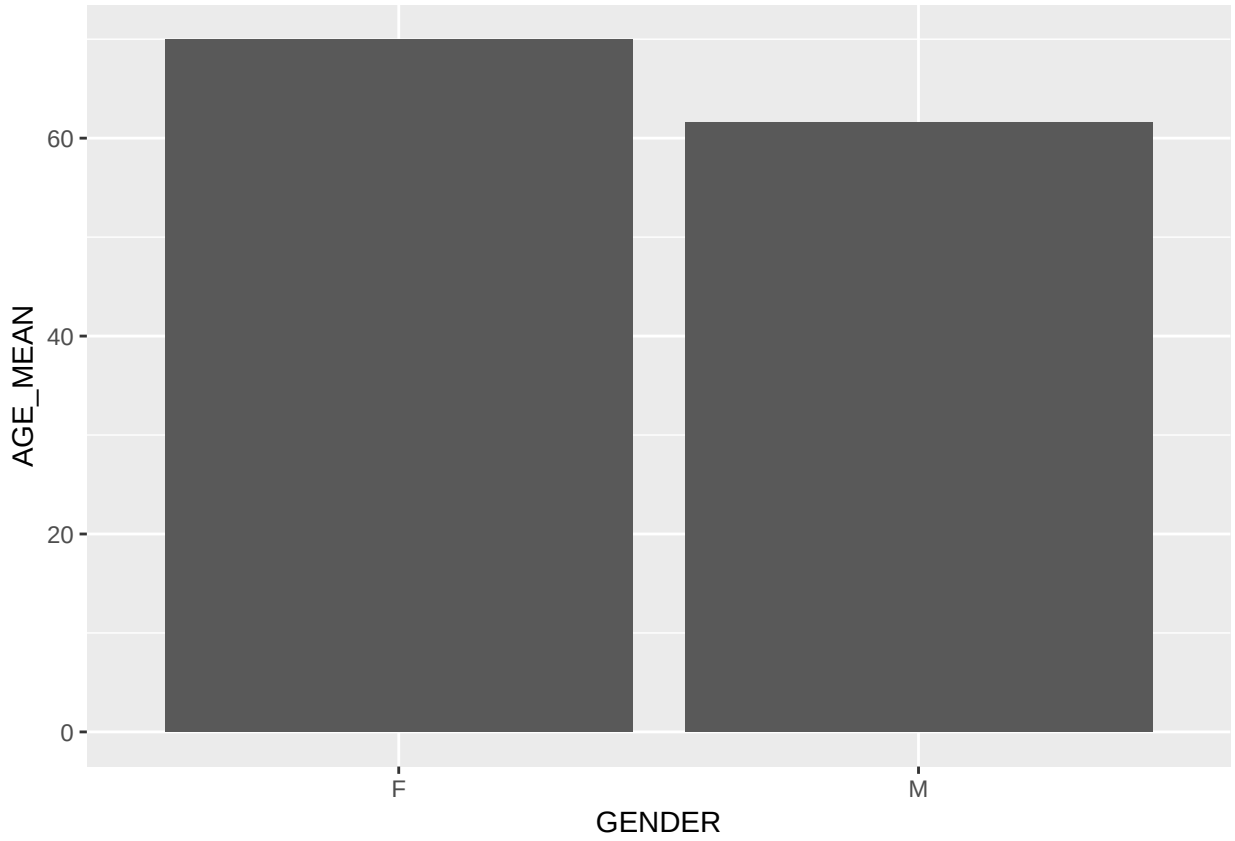

```
plot1 <- ggplot() +
  geom_histogram(data = human,
                 aes(x = AGE),
                 binwidth = 1)

plot1
```



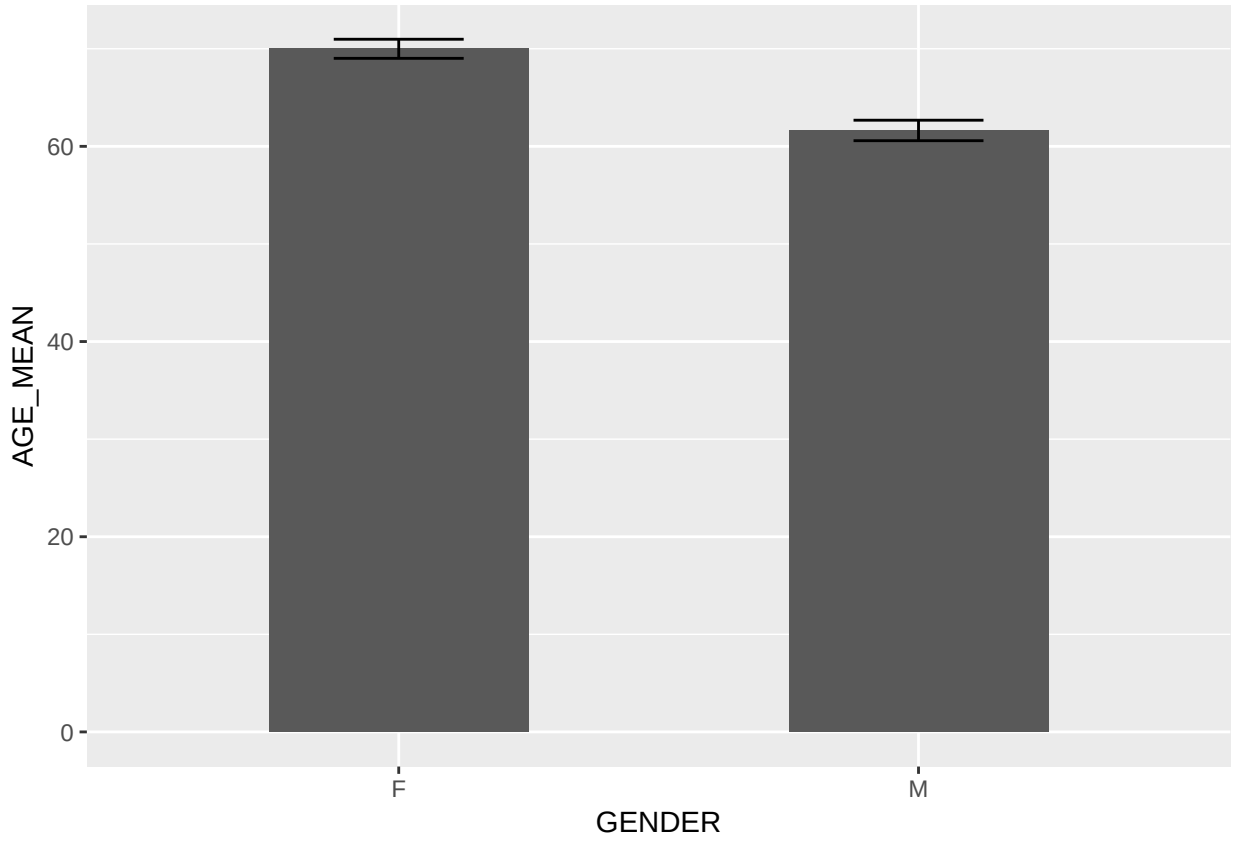

```
plot1 <- ggplot() +
  geom_histogram(data = human,
                 aes(x = AGE, fill = GENDER),
                 binwidth = 1)

plot1
```


```
human_avg <- human %>%
  group_by(GENDER) %>%
  summarise(AGE_MEAN = mean(AGE),
            AGE_SE = sd(AGE)/sqrt(n()))
```

```
plot2 <- ggplot() +
  geom_col(data = human_avg,
           aes(x = GENDER, y = AGE_MEAN))

plot2
```




```
plot2 <- ggplot() +
  geom_col(data = human_avg,
           aes(x = GENDER, y = AGE_MEAN),
           width = 0.5) +
  geom_errorbar(data = human_avg,
                aes(x = GENDER, ymin = AGE_MEAN - AGE_SE, ymax = AGE_MEAN + AGE_SE),
                width = 0.25)

plot2
```




```
plot3 <- ggplot() +
  geom_col(data = human_avg,
           aes(x = GENDER, y = AGE_MEAN),
           width = 0.5) +
  geom_errorbar(data = human_avg,
                aes(x = GENDER, ymin = AGE_MEAN - AGE_SE, ymax = AGE_MEAN + AGE_SE),
                width = 0.25) +
  xlab("Gender") +
  ylab("Age at death (mean)")

plot3
```

Statistical analysis

```
male <- human %>%
  filter(GENDER == "M") # filters only males out to assign to 'male'

female <- human %>%
  filter(GENDER == "F") # filters only females out to assign to 'female'
```

```
human_ttest <- t.test(male$AGE, female$AGE)

human_ttest
```

```
##
```

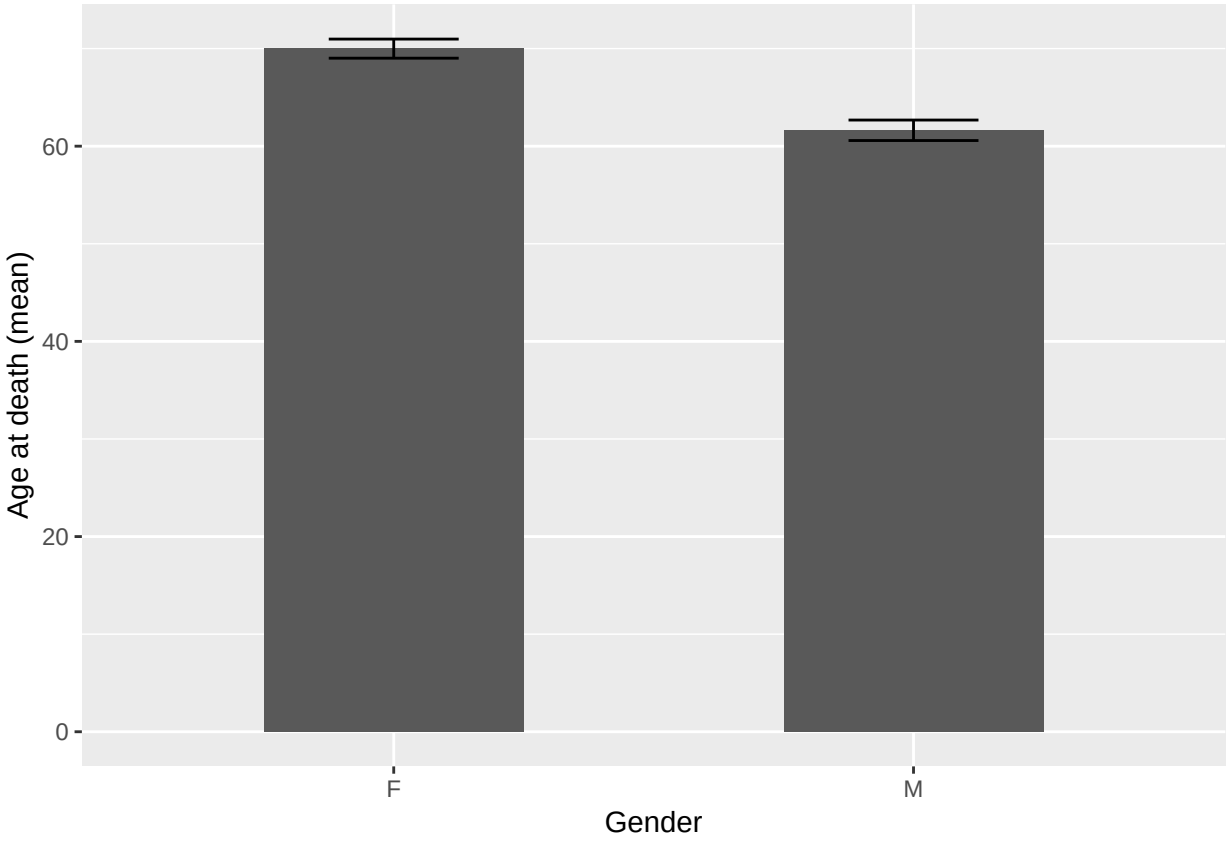


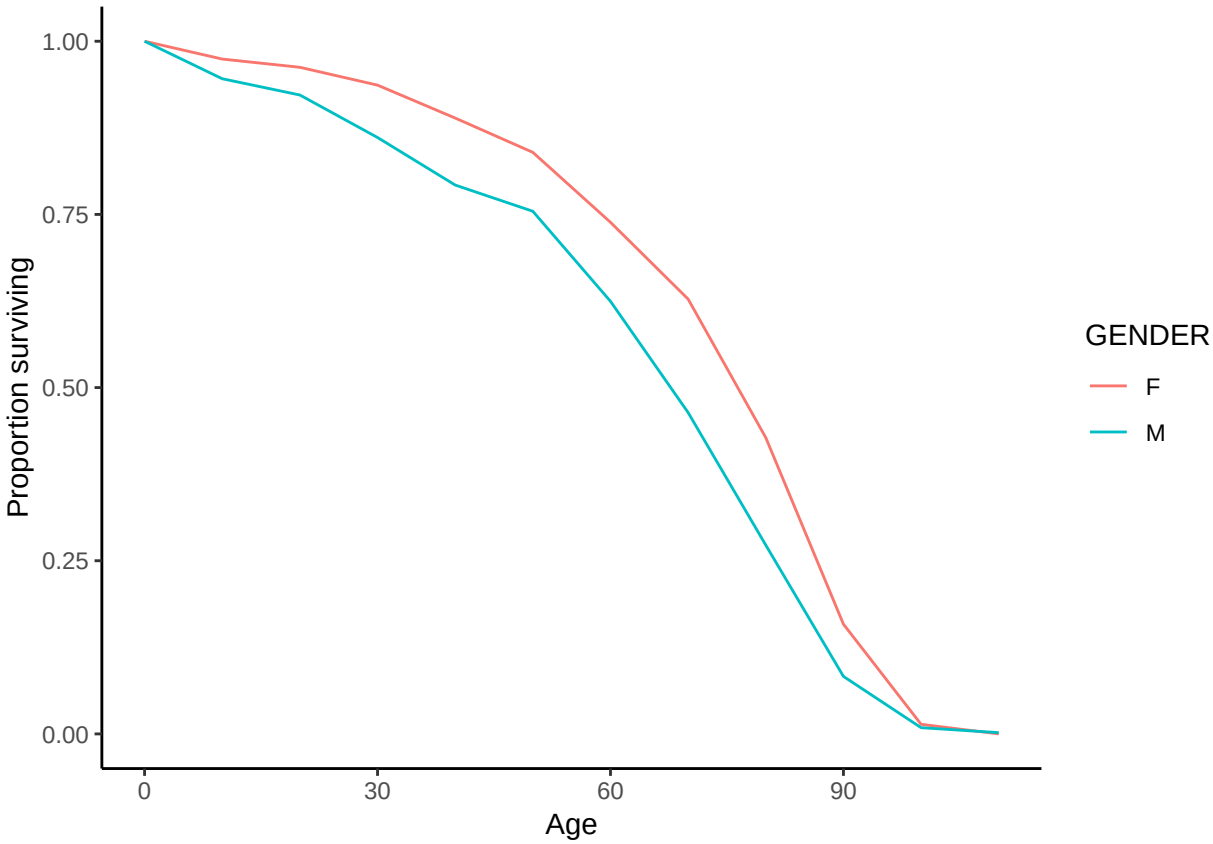
Figure 1: This is a really nice way to make figure captions without worrying about formatting in Microsoft Word

```
##  Welch Two Sample t-test
##
## data:  male$AGE and female$AGE
## t = -5.8325, df = 1056.4, p-value = 7.252e-09
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.18065  -5.55150
## sample estimates:
## mean of x mean of y
##  61.62996  69.99604
```

Survivorship curve

```
human_survivorship <- human %>%
  select(GENDER, AGE) %>%
  group_by(GENDER) %>%
  summarise(greater_0 = sum(AGE>=0)/n(),
            greater_10 = sum(AGE >=10)/n(),
            greater_20 = sum(AGE >=20)/n(),
            greater_30 = sum(AGE >=30)/n(),
            greater_40 = sum(AGE >= 40)/n(),
            greater_50 = sum(AGE >= 50)/n(),
            greater_60 = sum(AGE >= 60)/n(),
            greater_70 = sum(AGE >=70)/n(),
            greater_80 = sum(AGE >= 80)/n(),
            greater_90 = sum(AGE >= 90)/n(),
            greater_100 = sum(AGE >= 100)/n(),
            greater_110 = sum(AGE >= 110)/n()) %>%
  group_by(GENDER) %>%
  gather("AGE_RANGE", "COUNT", 2:ncol(.)) %>%
  mutate(AGE_RANGE = as.numeric(str_extract(AGE_RANGE, "[[:digit:]]+")))
```

```
plot4 <- ggplot() +
  geom_line(data = human_survivorship,
            aes(x = AGE_RANGE, y = COUNT, group = GENDER, colour = GENDER)) +
  theme_classic() +
  xlab("Age") +
  ylab("Proportion surviving")

plot4
```


1.00
0.75
0.50
0.25
0.00
Proportion surviving
0
30
60
90
Age
GENDER
F
M