

# Exercise 2 - Visa Network Data

*Lisa Reiber*

*20 11 2017*

```
#Get Data
nobel <- read.csv(file = "data/nobel.csv")
#Load Packages
if (!require("pacman")) install.packages("pacman")
pacman::p_load(tidyverse, statar)
```

## Exercise 1a

What is the distribution of females/males by category of the Nobel Prize (absolute and in percentages)?

**Answer 1a:**

```
statar::tab(nobel, gender)
```

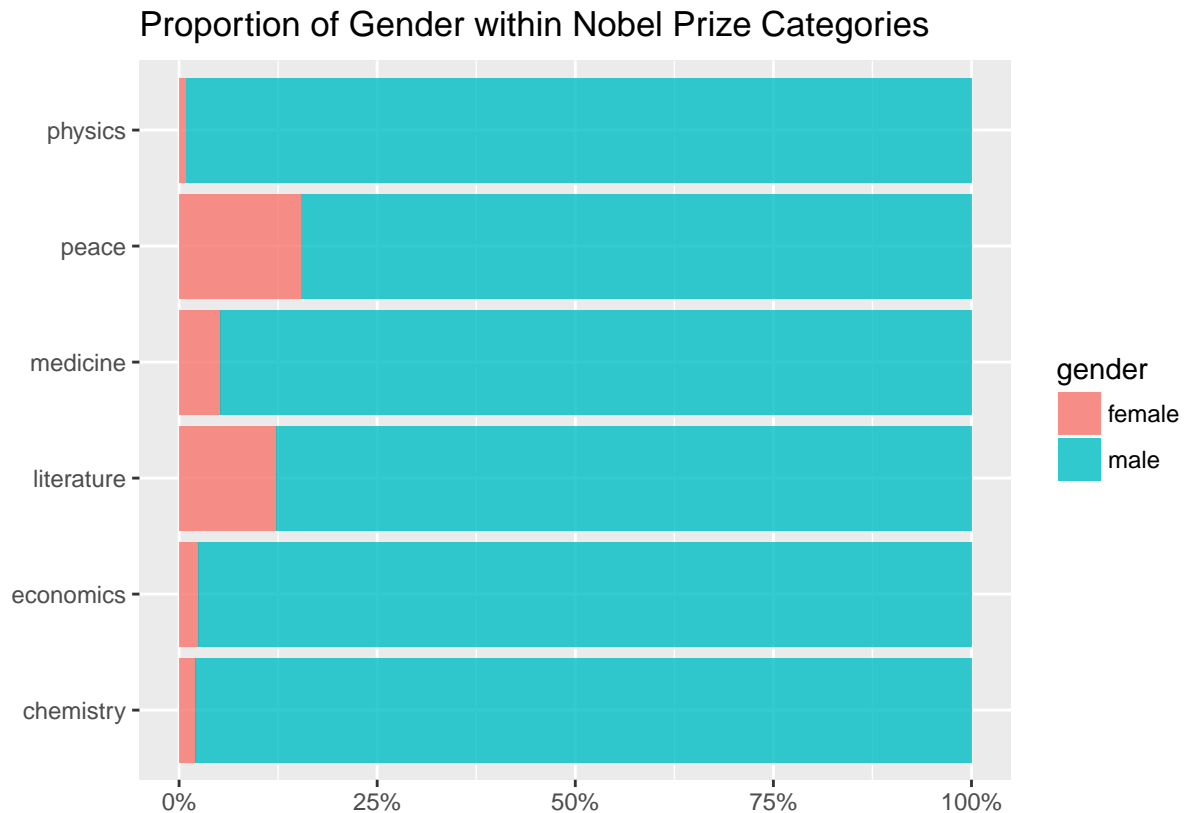
```
##
## gender      Freq.  Percent    Cum.
##
## female        50     5.05     5.05
##   male       913    92.22    97.27
##    org        27     2.73   100.00
```

## Exercise 1b: Please create a visualization using ggplot2.

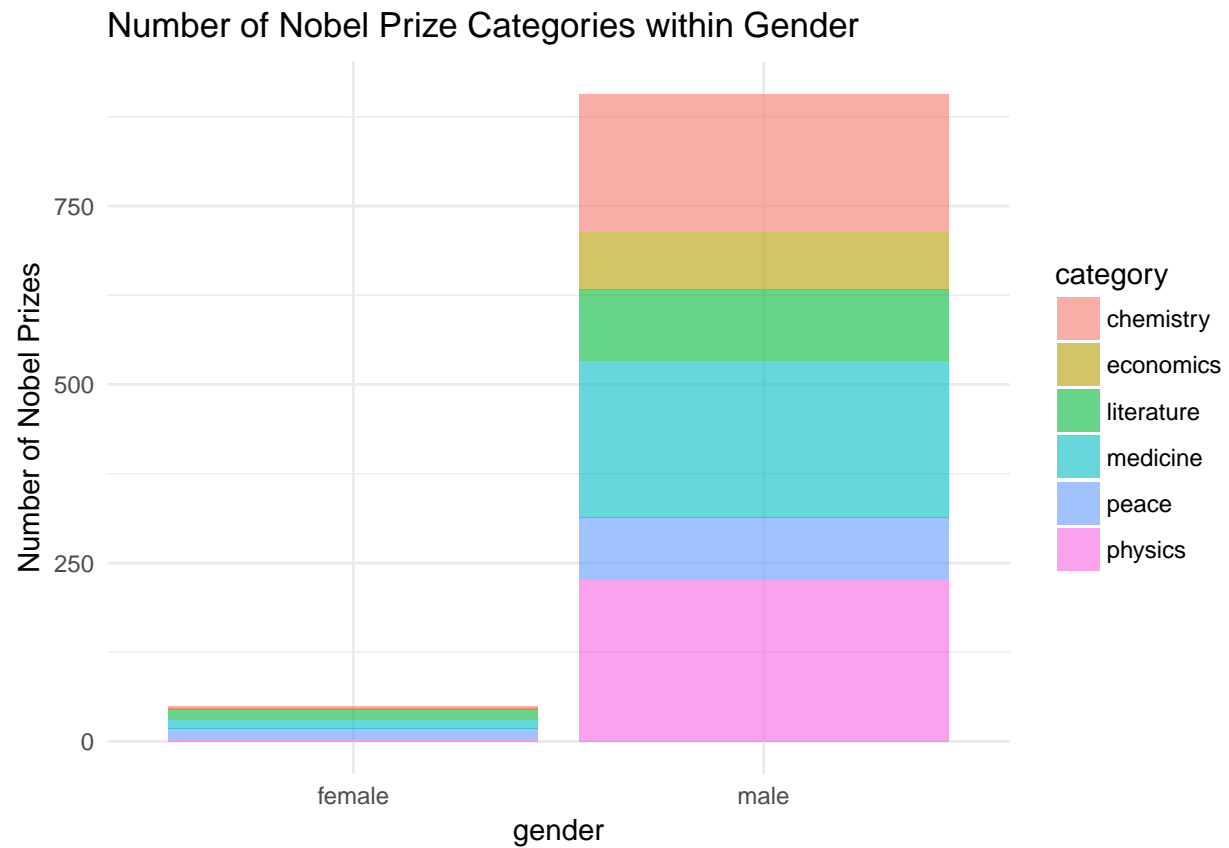
Answer 1b:

```
# prep
nobel_plot <- nobel %>% filter(gender != "org", category != "")

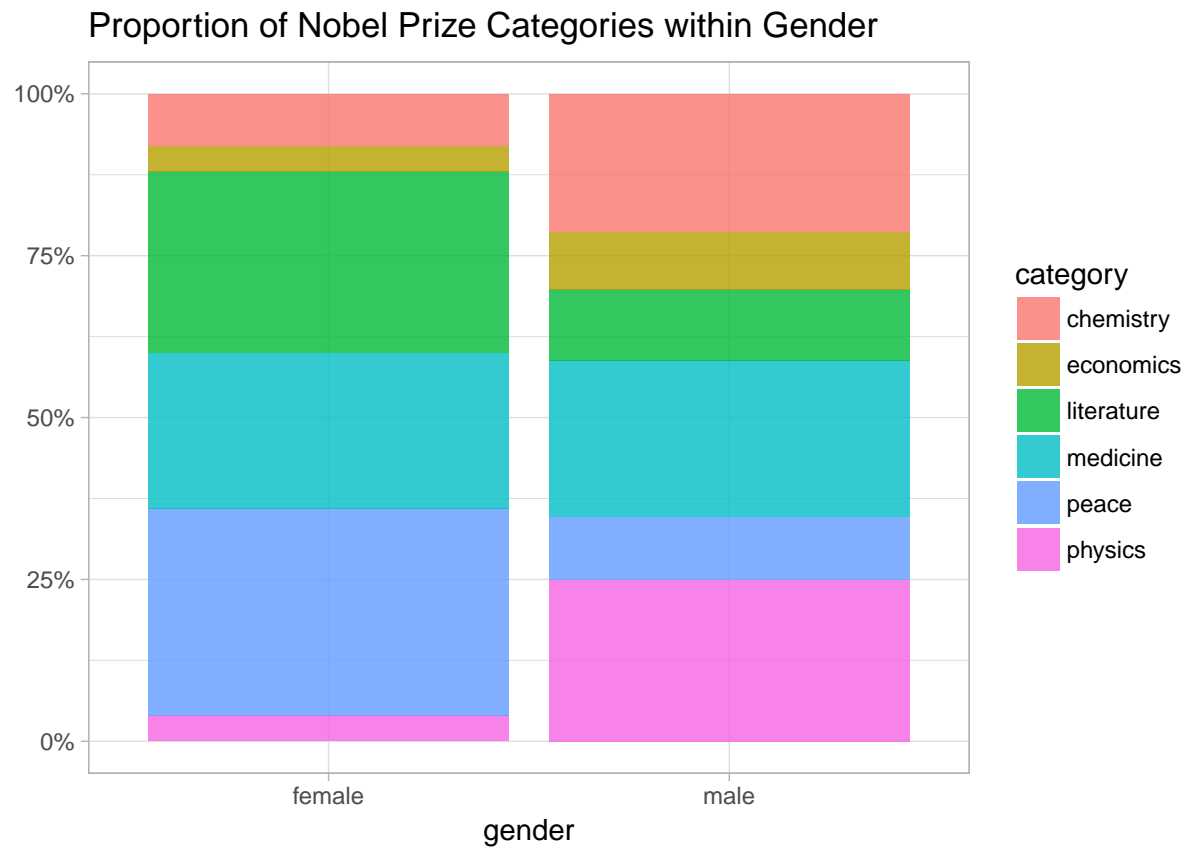
# 1. Horizontal Bar Chart, Categories by Gender (Proportion)
nobel_plot %>%
  ggplot(aes(category, fill = gender)) +
  geom_bar(alpha = 4/5, position = position_fill(reverse = TRUE)) +
  coord_flip() +
  scale_y_continuous(labels = scales::percent) +
  labs(title = "Proportion of Gender within Nobel Prize Categories",
       y = "",
       x = "")
```



```
# visualize inequality
# 2. Vertical Bar Chart, Gender by Categories (Count)
nobel_plot %>%
  ggplot(aes(gender, fill = category)) +
  geom_bar(alpha = 3/5, position = "stack") +
  theme_minimal() +
  labs(title = "Number of Nobel Prize Categories within Gender",
        y = "Number of Nobel Prizes")
```

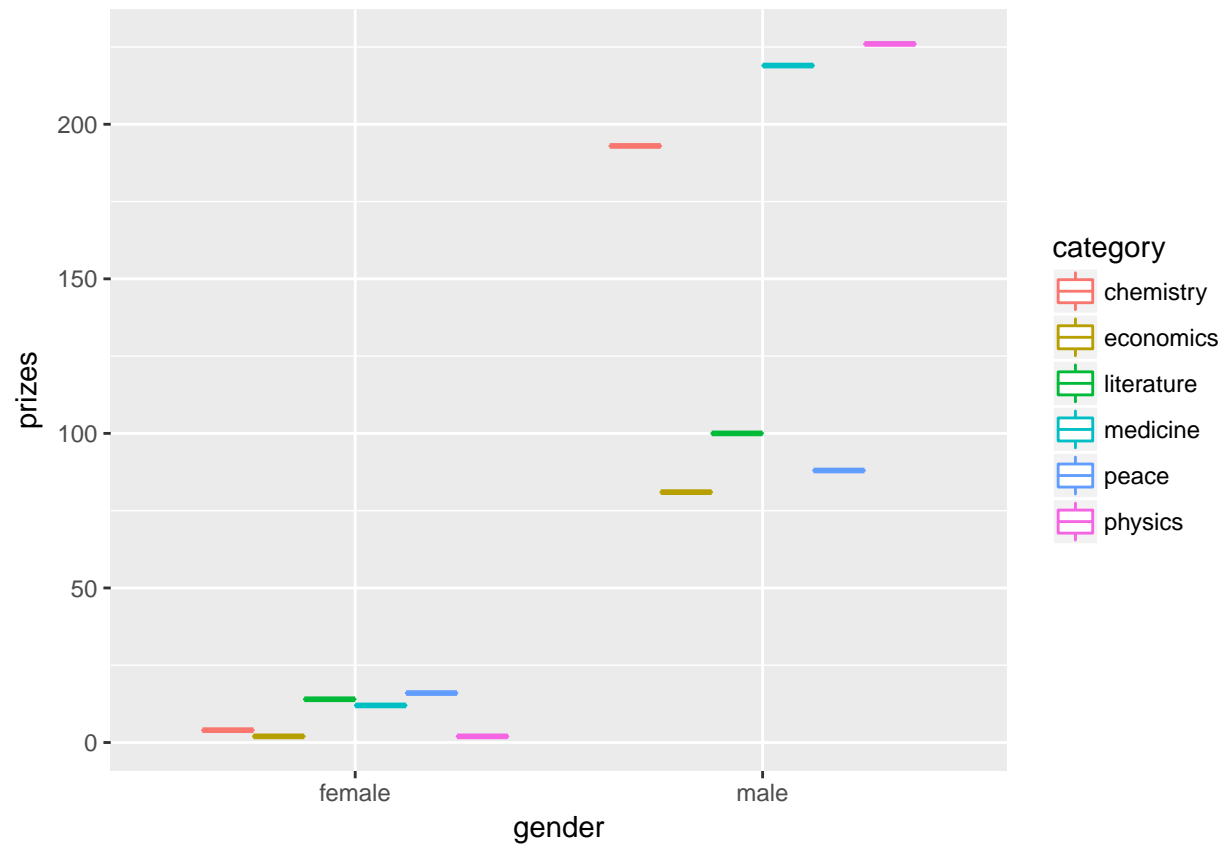


```
# visualize both distributions, hide inequality/ highlight distrib of categ.
# 3. Vertical Bar Chart, Gender by Categories (Proportion)
nobel_plot %>%
  ggplot(aes(gender, fill = category)) +
  geom_bar(alpha = 4/5, position = "fill") +
  theme_light() +
  scale_y_continuous(labels = scales::percent) +
  labs(title = "Proportion of Nobel Prize Categories within Gender",
       y = "")
```



```
# 4. boxplot number of prizes
```

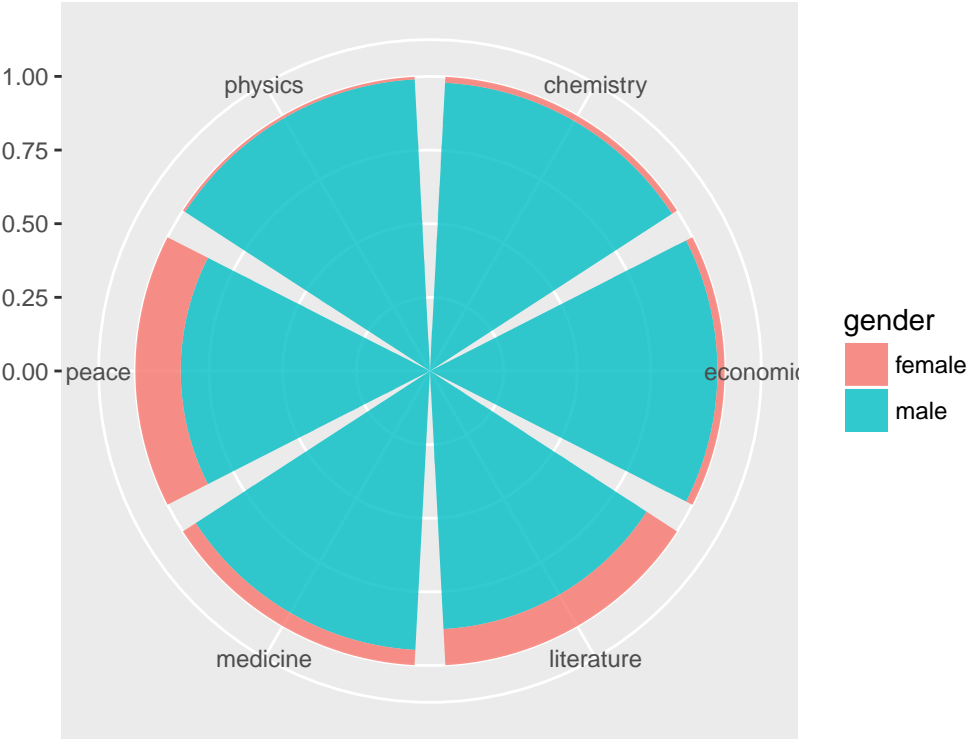
```
nobel_plot %>%
  group_by(gender, category) %>%
  summarize(prizes = n()) %>%
  ggplot(aes(x= gender, y= prizes, colour = category)) +
  geom_boxplot()
```



```
# 5.. hahaha
```

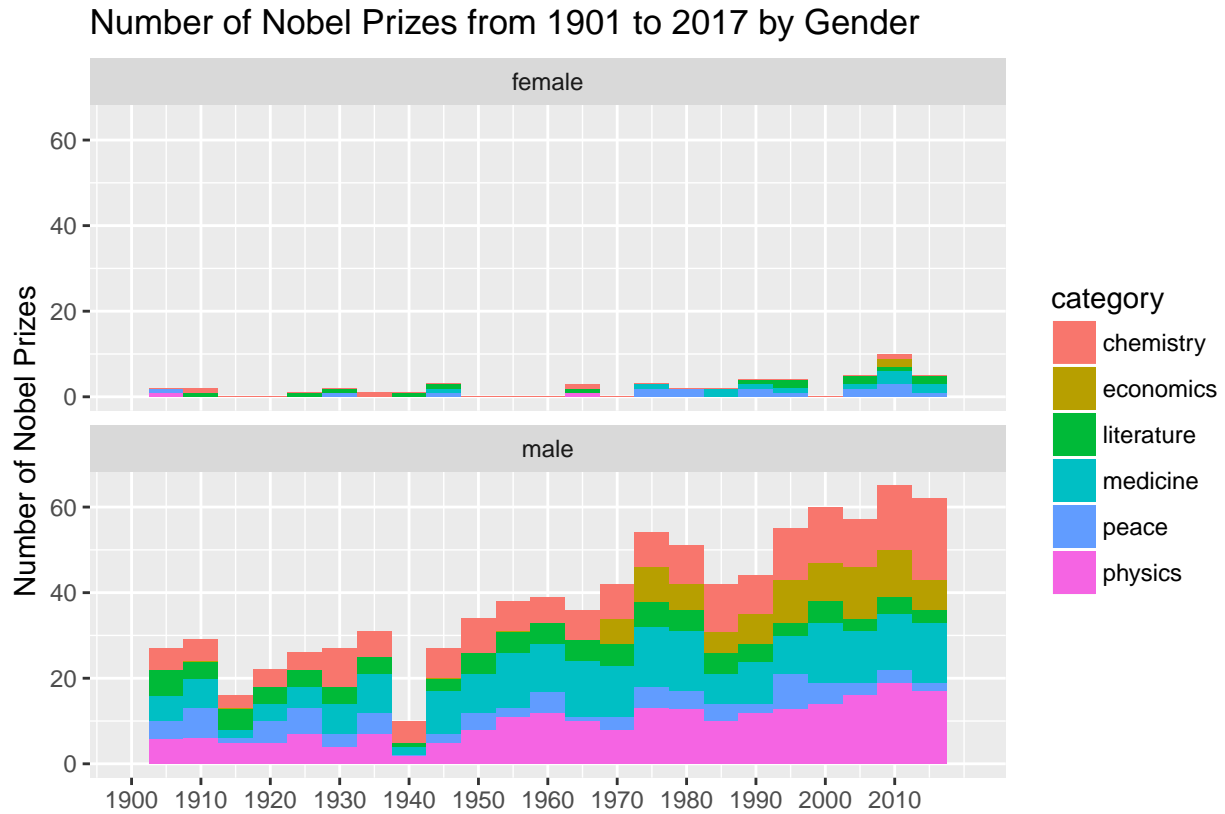
```
nobel_plot %>%
  ggplot(aes(category, fill = gender)) +
  geom_bar(alpha = 4/5, position = position_fill()) +
  coord_polar() +
  labs(title = "Proportion of Gender within Nobel Prize Categories",
       y = "",
       x="")
```

Proportion of Gender within Nobel Prize Categories



# 5. Nobel Prizes by year over gender

```
ggplot(data = nobel_plot, mapping = aes(x = year, fill = category)) +
  geom_histogram(position = "stack", binwidth = 5) +
  facet_wrap(c("gender"), scales = , dir = "v") +
  scale_x_continuous(breaks = seq(1900, 2010, 10), limits = c(1900, 2020)) +
  labs(title = "Number of Nobel Prizes from 1901 to 2017 by Gender ",
       y = "Number of Nobel Prizes", x = "")
```



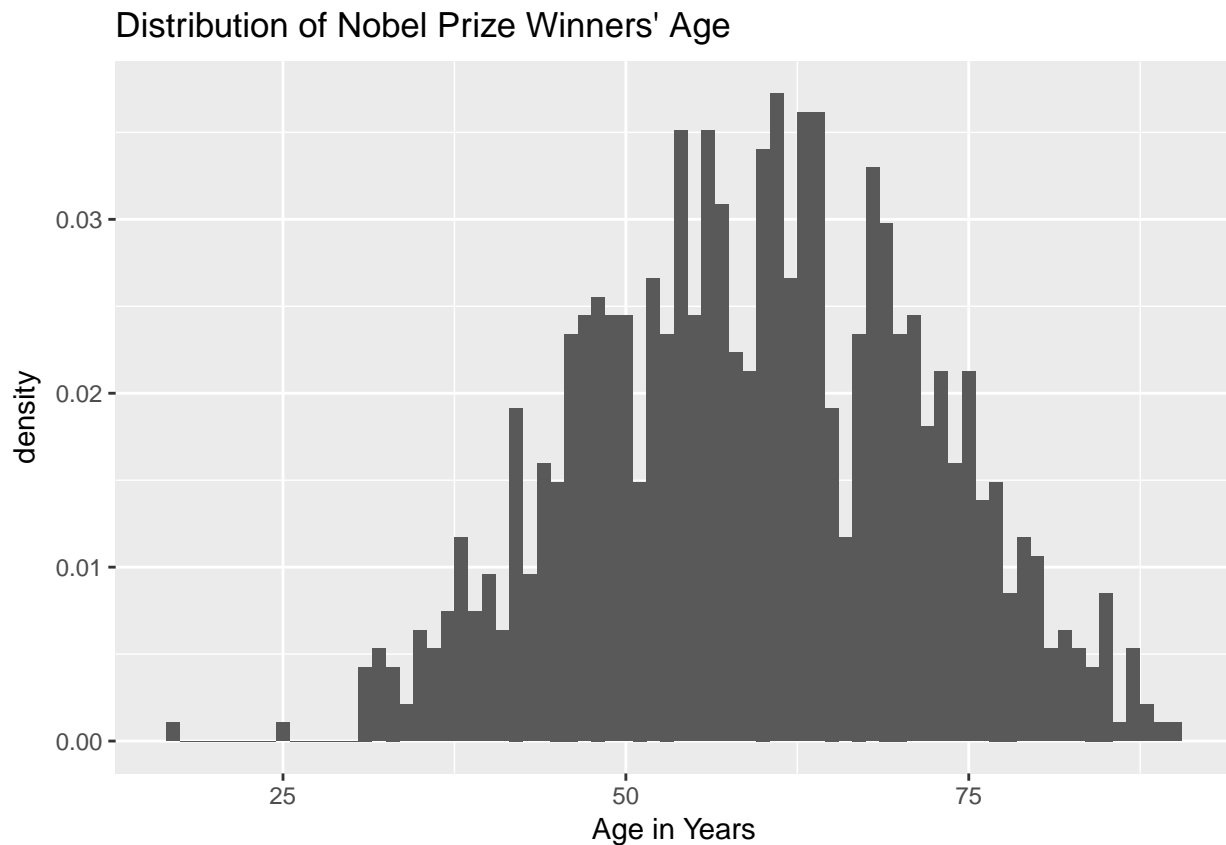
## Exercise 2:

How does the age distribution of Nobel Prize Laureates look like (their age when they received the prize)? Again, please visualize your result.

### Answer 2:

```
#prep
library(stringr)
nobel_plot <- nobel_plot %>%
  mutate(born = strtoi(str_extract_all(born, "[[:digit:]]{4}"))) %>%
  filter(born > 0) %>%
  mutate(age_win = year - born)

# histogram
nobel_plot %>%
  ggplot(aes(x= age_win, y= ..density..)) +
  geom_histogram(binwidth = 1) +
  labs(title = "Distribution of Nobel Prize Winners' Age",
       x = "Age in Years")
```



```
# boxplot
nobel_plot %>%
  ggplot(aes(x= gender, y= age_win, fill = category)) +
  geom_boxplot() +
```



```
labs(title = "Distribution of Nobel Prize Winners' Age by Categories and Gender",
     y = "Age in Years", x = "")
```

