

Time Series

Alan Arnholt

8/30/2016

Read in the data:

```
library(readr)
ages <- read_csv("Ch4Eg.csv")
ages
```

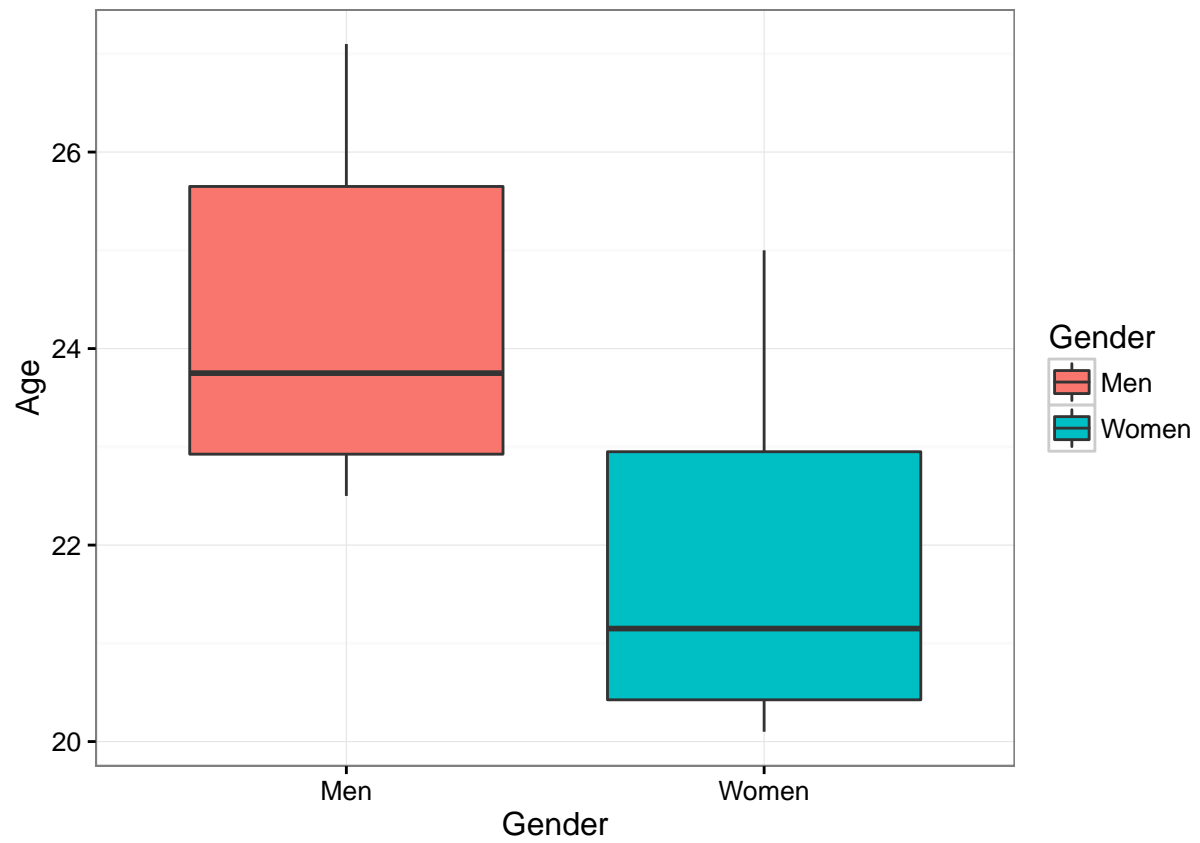
```
# A tibble: 58 × 3
  Year Women  Men
  <int> <dbl> <dbl>
1  1998  25.0  26.7
2  1997  25.0  26.8
3  1996  24.8  27.1
4  1995  24.5  26.9
5  1994  24.5  26.7
6  1993  24.5  26.5
7  1992  24.4  26.5
8  1991  24.1  26.3
9  1990  23.9  26.1
10 1989  23.8  26.2
# ... with 48 more rows
```

Create side-by-side boxplots:

```
library(tidyr)
NDF <- gather(ages, Gender, Age, -Year)
NDF
```

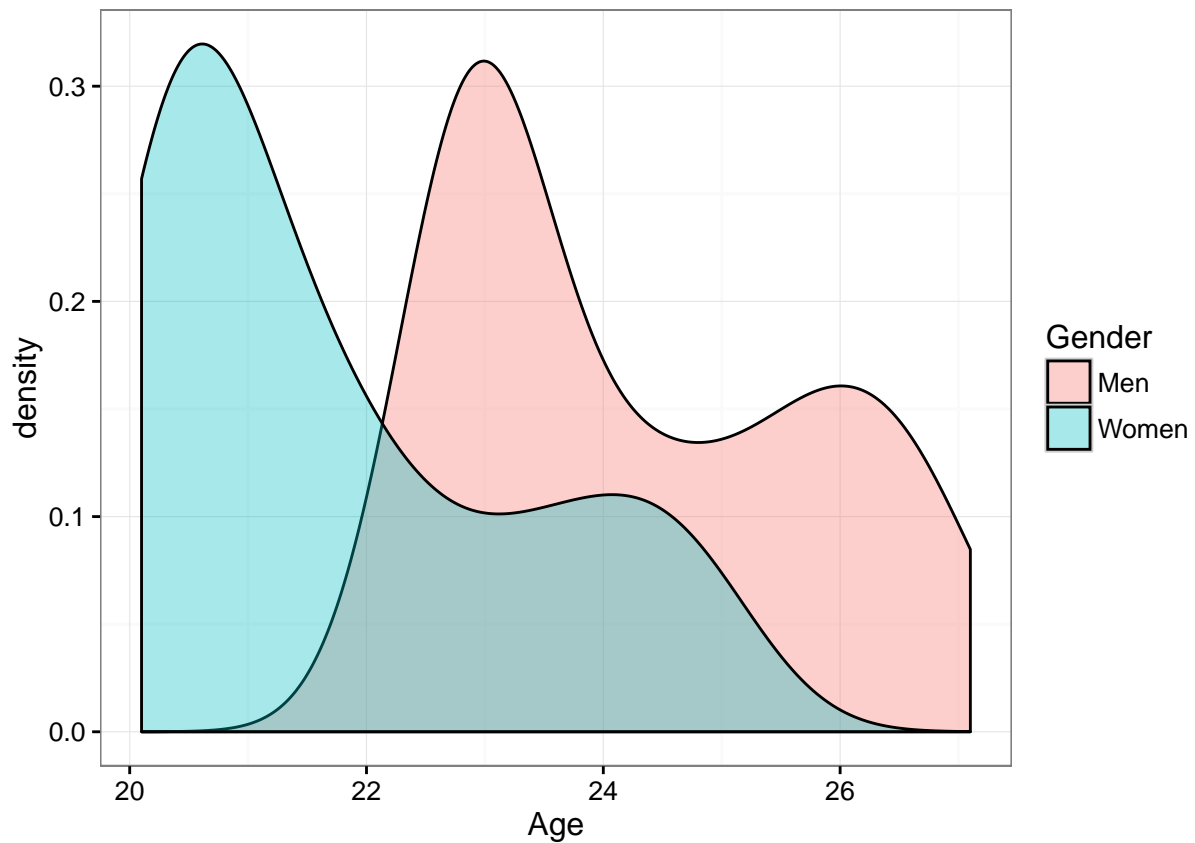
```
# A tibble: 116 × 3
  Year Gender  Age
  <int> <chr> <dbl>
1  1998 Women  25.0
2  1997 Women  25.0
3  1996 Women  24.8
4  1995 Women  24.5
5  1994 Women  24.5
6  1993 Women  24.5
7  1992 Women  24.4
8  1991 Women  24.1
9  1990 Women  23.9
10 1989 Women  23.8
# ... with 106 more rows
```

```
library(ggplot2)
ggplot(data = NDF, aes(x = Gender, y = Age, fill = Gender)) + geom_boxplot() +
  theme_bw()
```



Density plots:

```
ggplot(data = NDF, aes(x = Age, fill = Gender)) +  
  geom_density(alpha = 0.35) +  
  theme_bw()
```



Summary information:

```
library(dplyr)
SI <- NDF %>%
  group_by(Gender) %>%
  summarise(av_age = mean(Age), md_age = median(Age), sd_age = sd(Age))
SI
```

```
# A tibble: 2 × 4
  Gender av_age md_age sd_age
  <chr>   <dbl> <dbl> <dbl>
1 Men    24.25000 23.75 1.490820
2 Women  21.75172 21.15 1.554421
```

Create a time-series plot:

```
ggplot(data = ages, aes(x = Year, y = Women)) +
  geom_line(color = "purple") +
  geom_point(color = "purple") +
  geom_smooth(color = "pink") +
  geom_line(aes(x = Year, y = Men), color = "blue") +
  geom_point(aes(x = Year, y = Men), color = "blue") +
  geom_smooth(aes(x = Year, y = Men), color = "lightblue") +
  labs(y = "Age") +
  theme_bw()
```

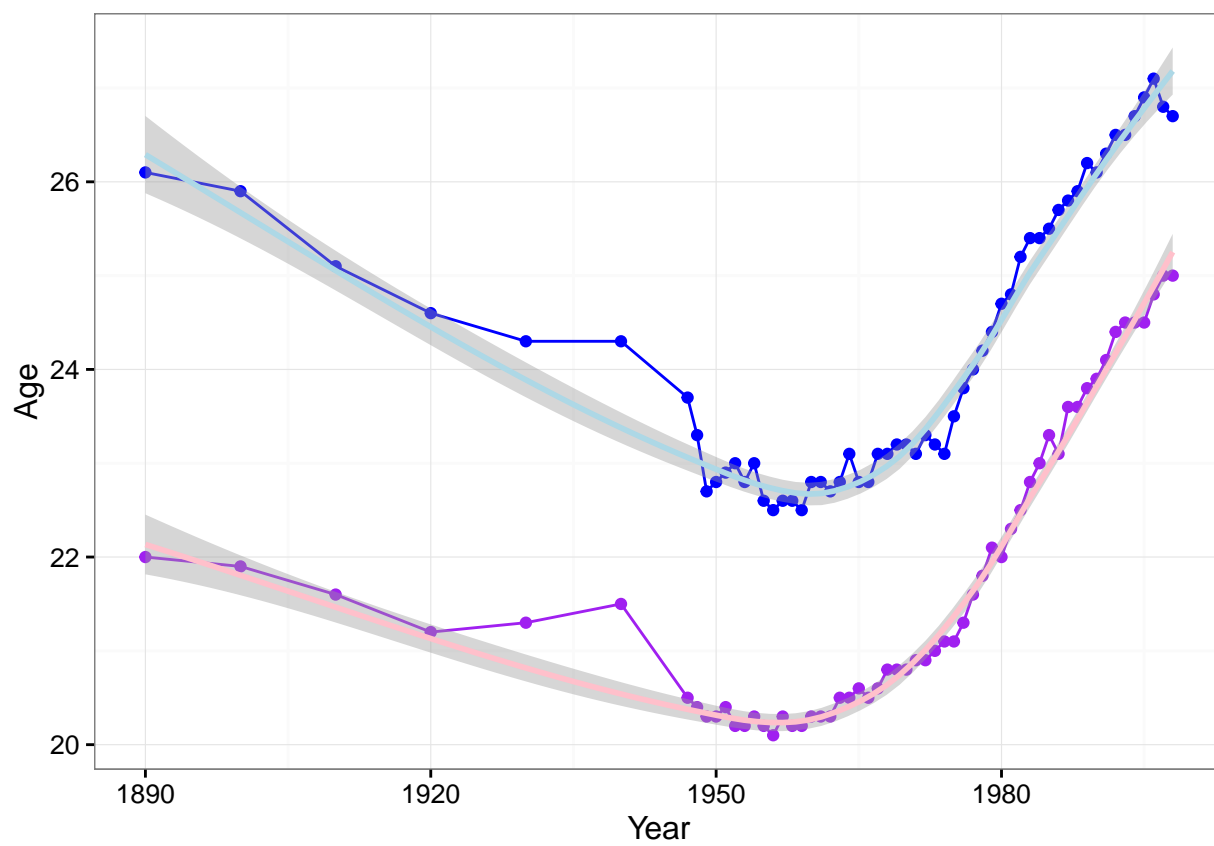


Figure 1: Time Series Plot

Another one using NDF:

```
ggplot(data = NDF, aes(x = Year, y = Age, color = Gender)) +  
  geom_point() +  
  geom_line() +  
  geom_smooth() +  
  theme_bw()
```

