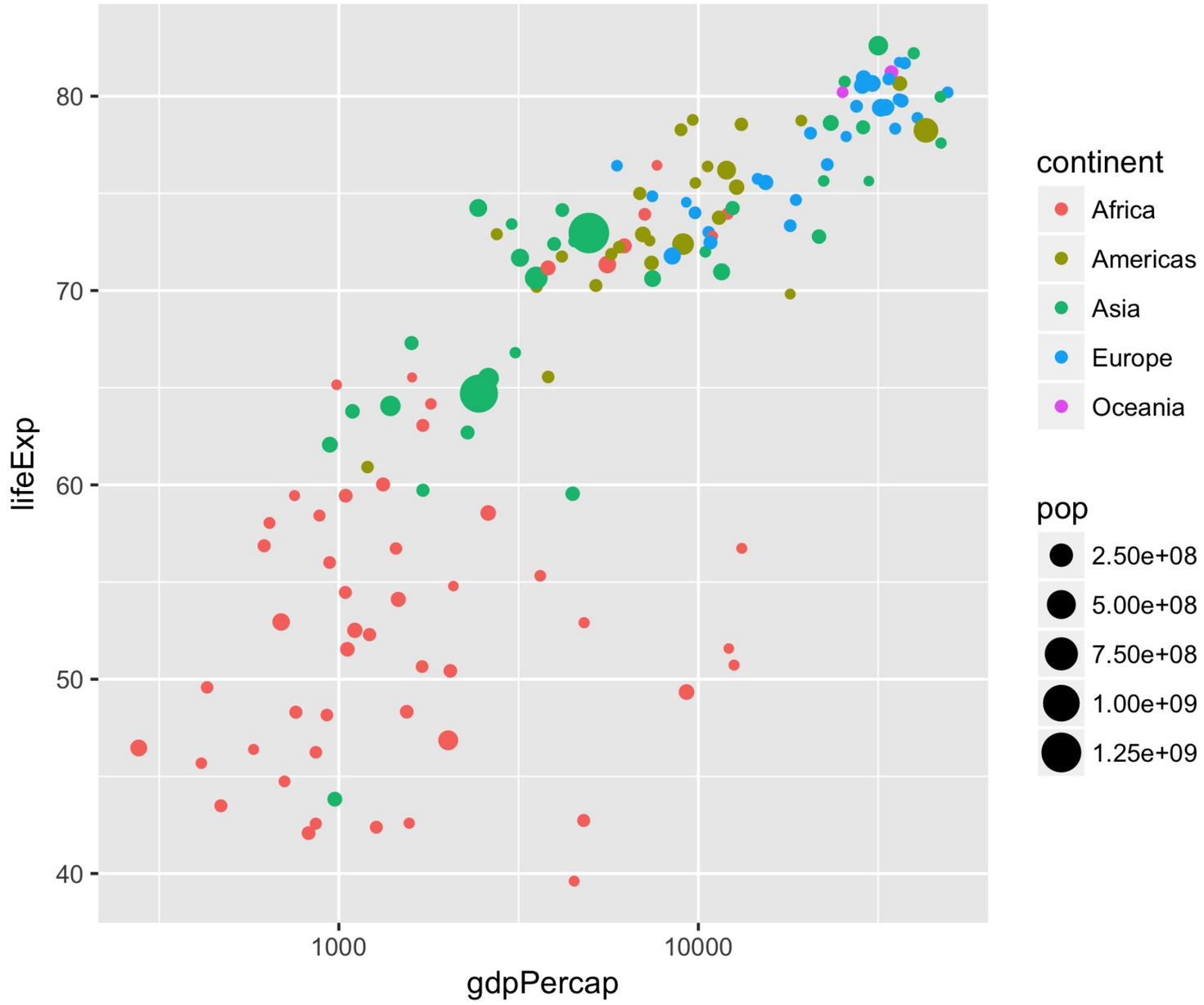
INTRODUCTION TO THE TIDYVERSE

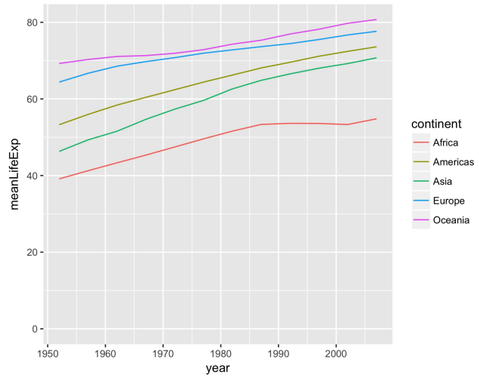
Line plots

David Robinson
Chief Data Scientist, DataCamp

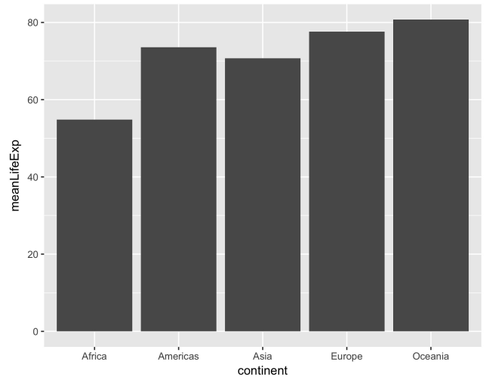

80
70
60
50
40
lifeExp
1000
10000
gdpPercap
continent
Africa
Americas
Asia
Europe
Oceania
pop
2.50e+08
5.00e+08
7.50e+08
1.00e+09
1.25e+09

Types of plots

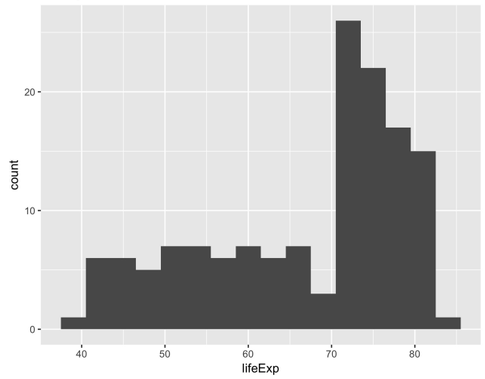
Line plot



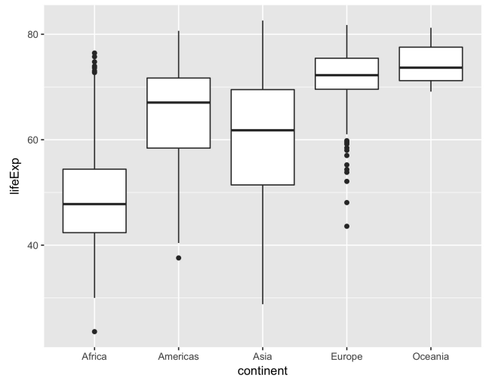
Bar plot



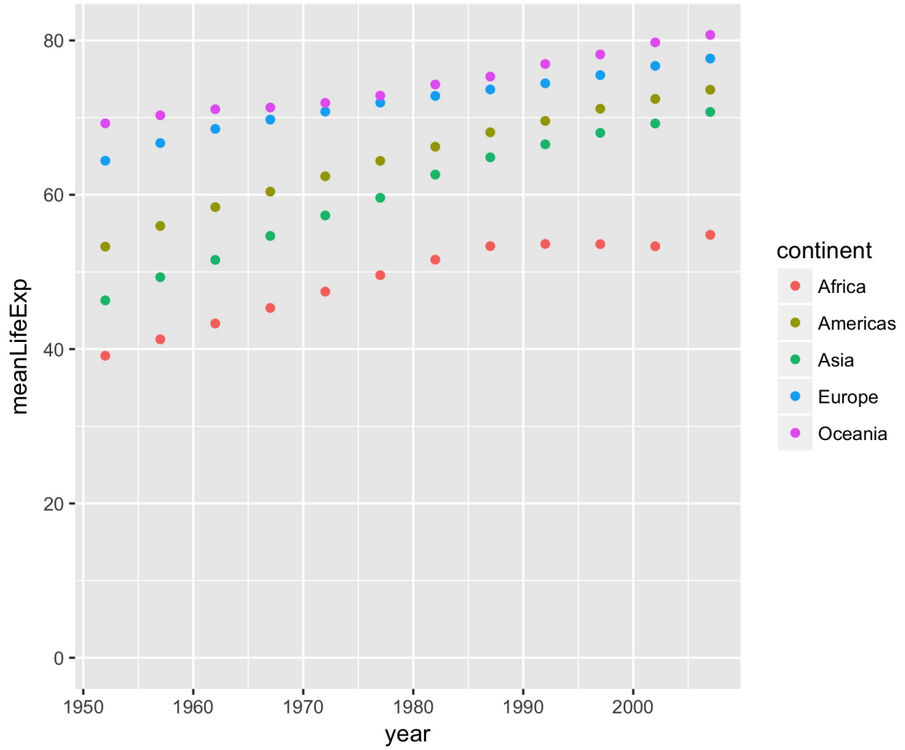
Histogram



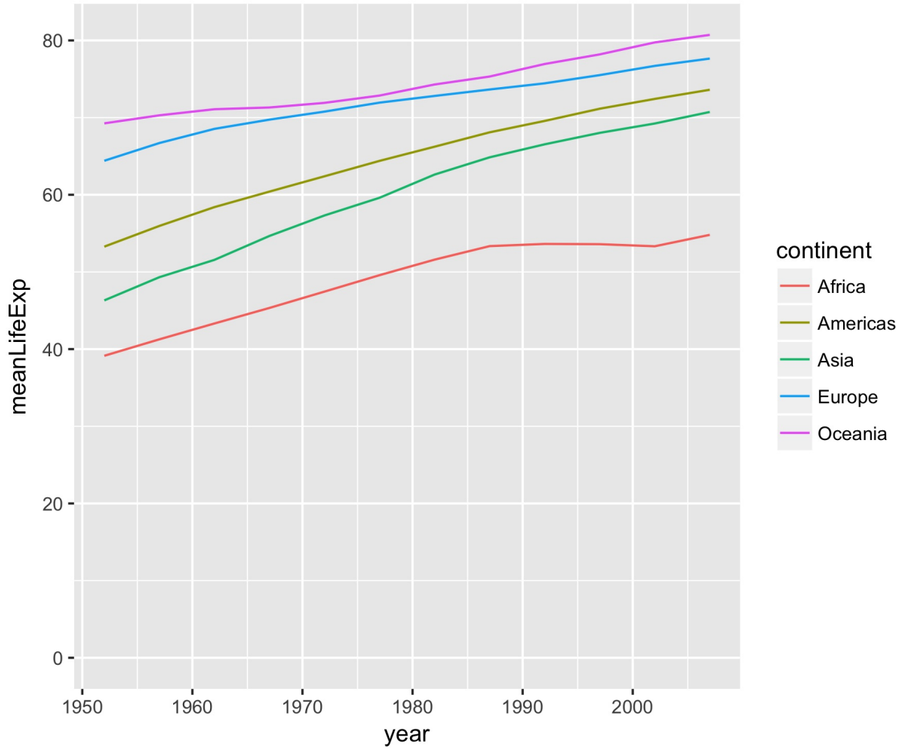
Box plot



Scatter vs line plot

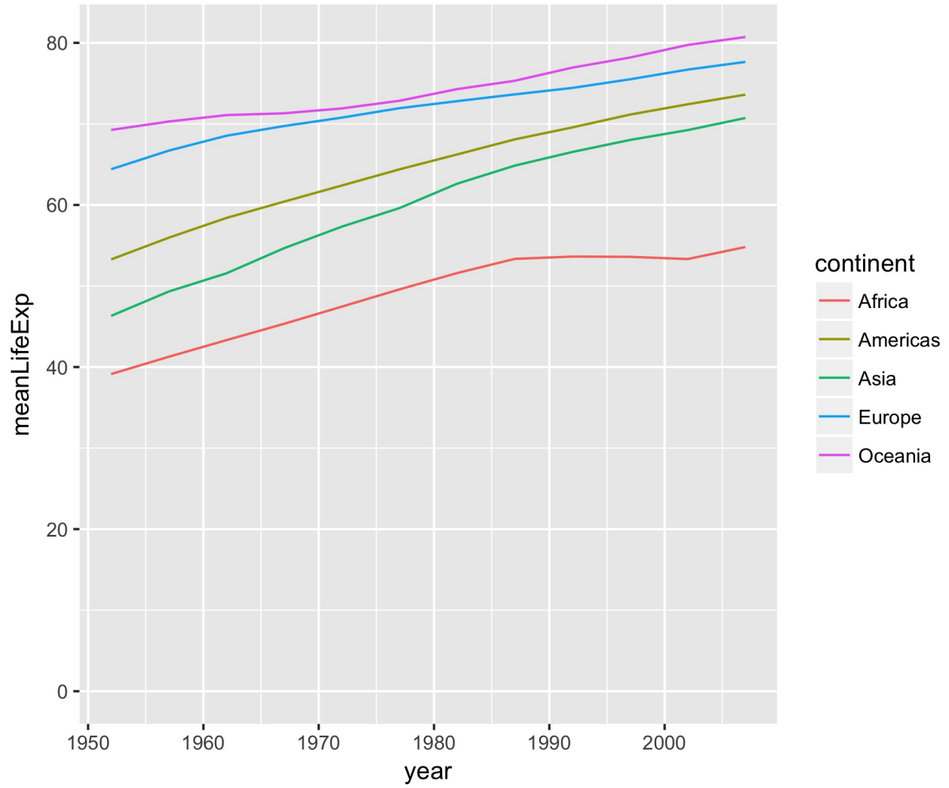


geom_point()



geom_line()

Line plot




```
ggplot(year_continent, aes(x = year, y = meanLifeExp, color = continent)) +
  geom_line() +
  expand_limits(y = 0)
```

INTRODUCTION TO THE TIDYVERSE

Let's practice!

INTRODUCTION TO THE TIDYVERSE

Bar plots

David Robinson
Chief Data Scientist, DataCamp

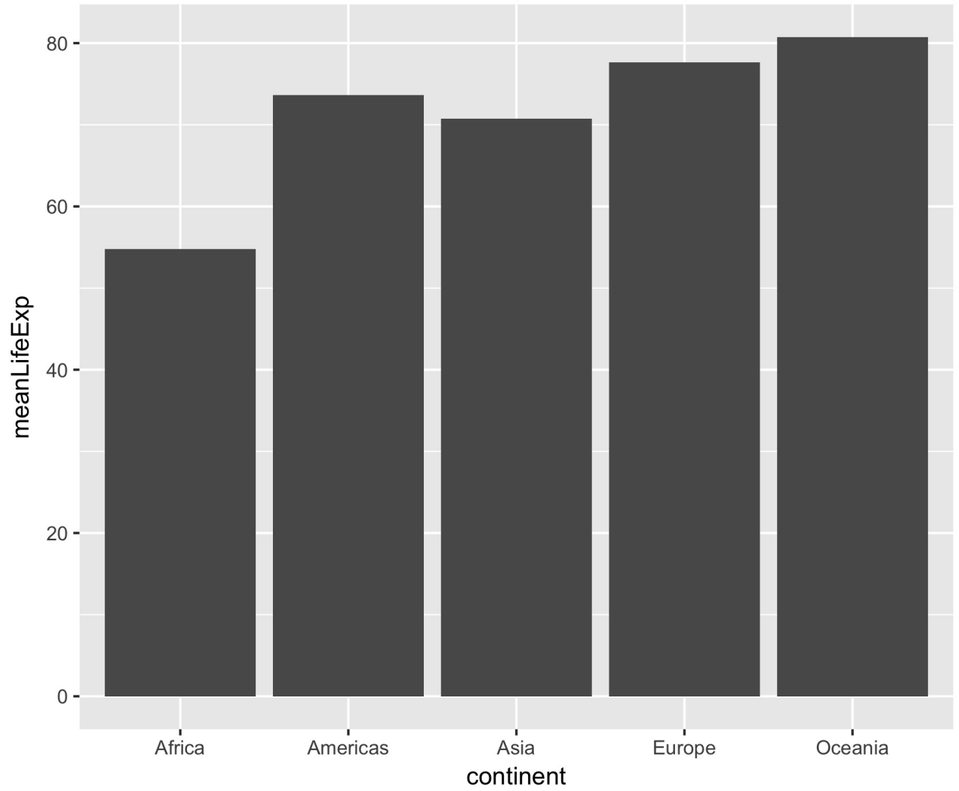
Summarizing by continent

```
by_continent <- gapminder %>%
  filter(year == 2007) %>%
  group_by(continent) %>%
  summarize(meanLifeExp = mean(lifeExp))

by_continent
```

```
# A tibble: 5 x 2
  continent meanLifeExp
     <fctr>       <dbl>
1    Africa    54.80604
2  Americas    73.60812
3      Asia    70.72848
4    Europe    77.64860
5   Oceania    80.71950
```

Bar plot




```
ggplot(by_continent, aes(x = continent, y = meanLifeExp)) +
  geom_col()
```

INTRODUCTION TO THE TIDYVERSE

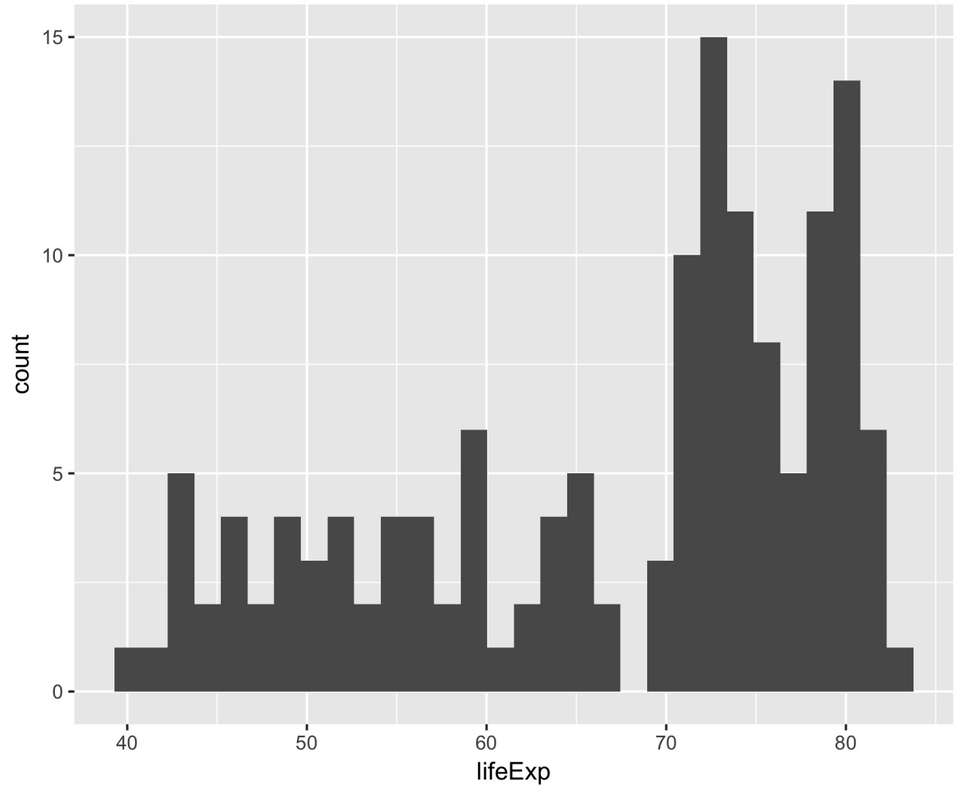
Let's practice!

INTRODUCTION TO THE TIDYVERSE

Histograms

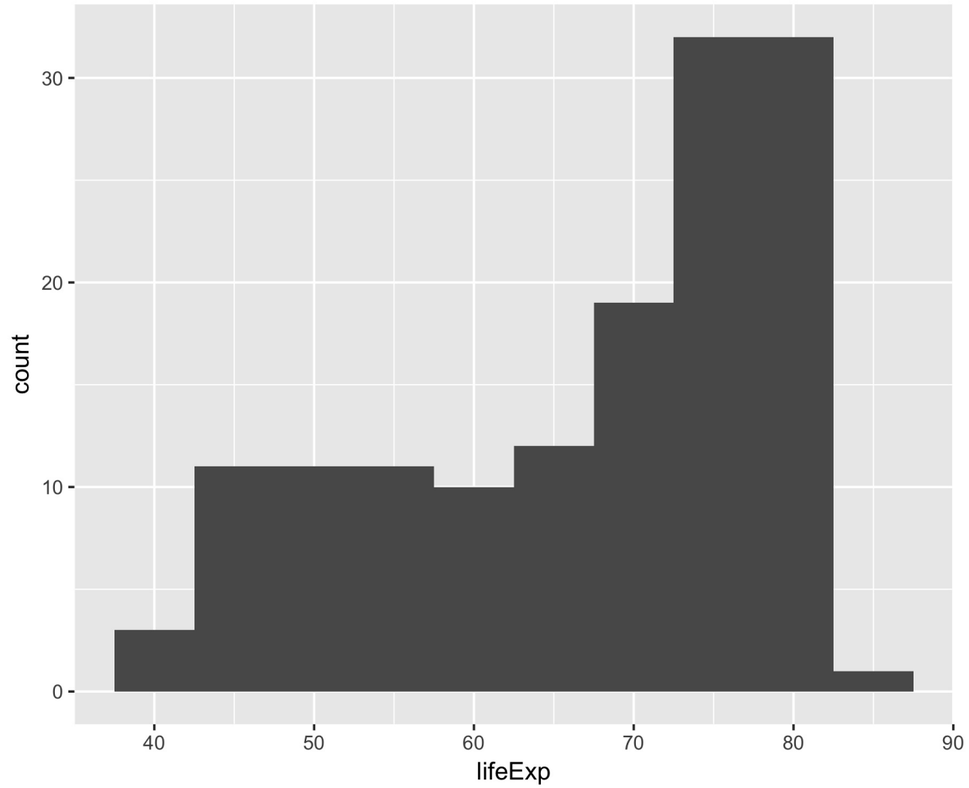
David Robinson
Chief Data Scientist, DataCamp

Histogram




```
ggplot(gapminder_2007, aes(x = lifeExp)) +
  geom_histogram()
```

Adjusting bin width




```
ggplot(gapminder_2007, aes(x = lifeExp)) +
  geom_histogram(binwidth = 5)
```

Log x-axis

```
scale_x_log10()
```

INTRODUCTION TO THE TIDYVERSE

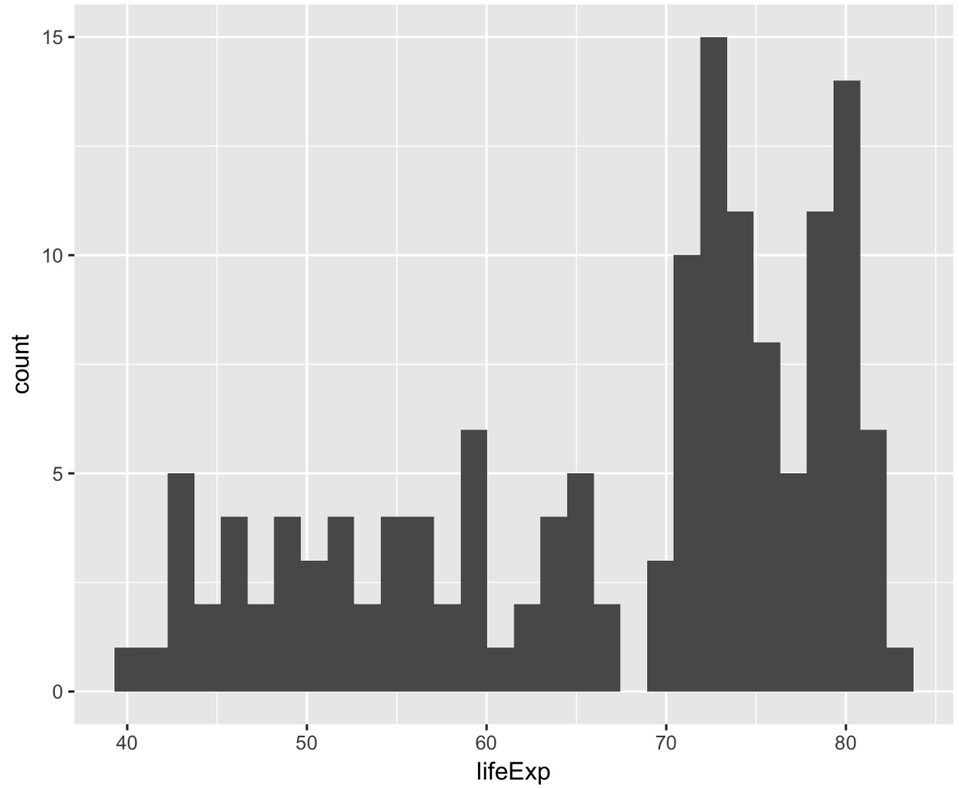
Let's practice!

INTRODUCTION TO THE TIDYVERSE

Box plots

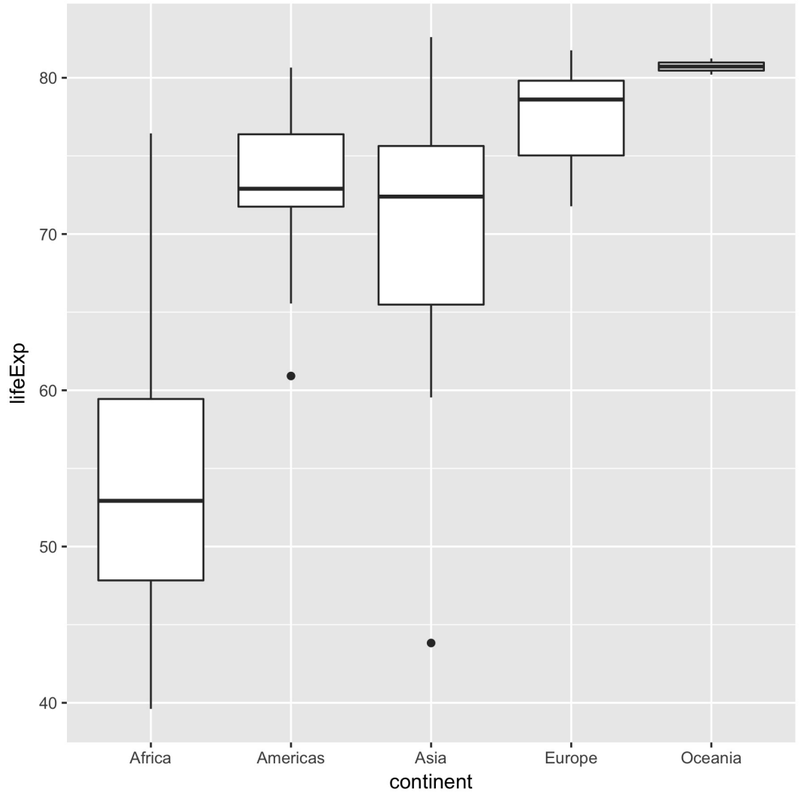
David Robinson
Chief Data Scientist, DataCamp

Histograms



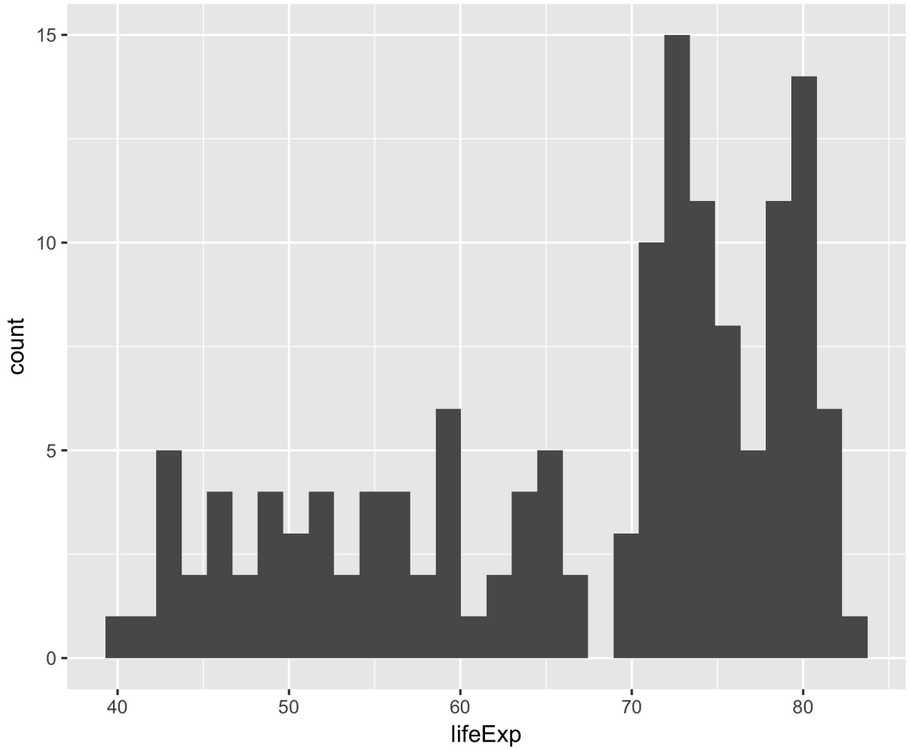

```
ggplot(gapminder_2007, aes(x = lifeExp)) +
  geom_histogram()
```

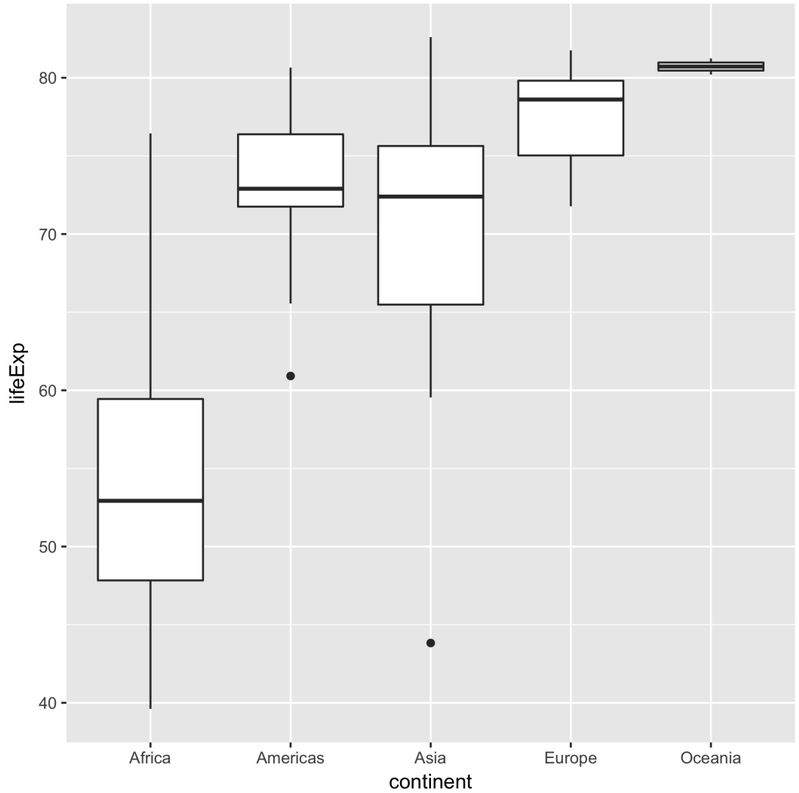
Box plots




```
ggplot(gapminder_2007, aes(x = continent, y = lifeExp)) +
  geom_boxplot()
```

Histogram vs box plot





INTRODUCTION TO THE TIDYVERSE

Let's practice!

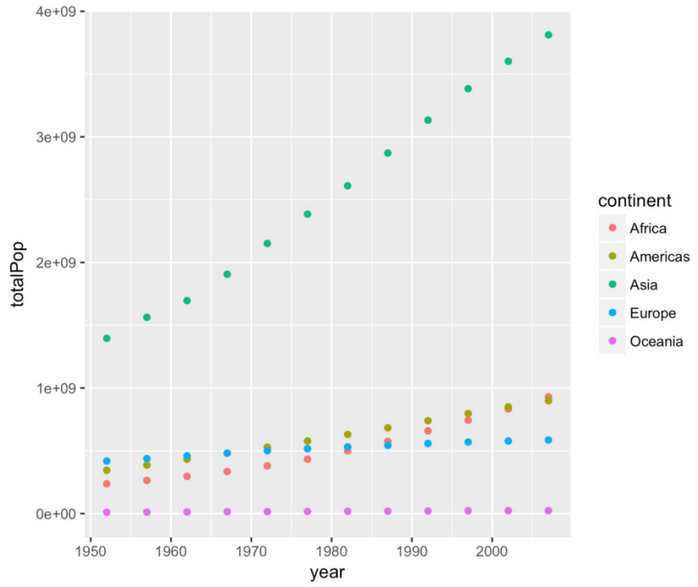
INTRODUCTION TO THE TIDYVERSE

Conclusion

David Robinson
Chief Data Scientist, DataCamp

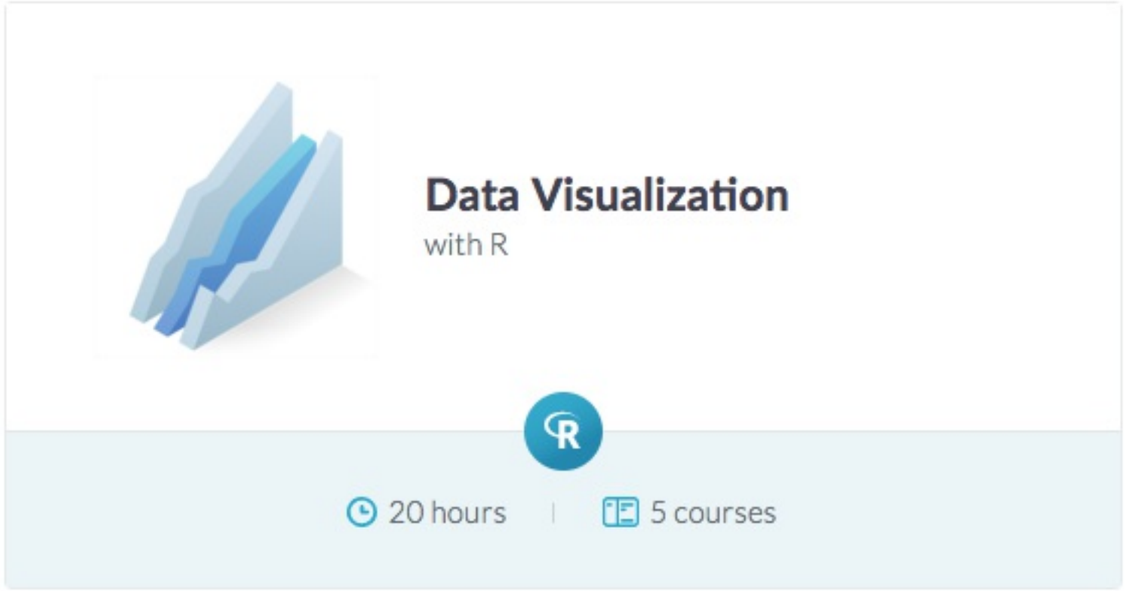
Transforming and visualizing data with R

```
ggplot(by_year_continent, aes(x = year, y = totalPop, color = continent)) +
  geom_point() +
  expand_limits(y = 0)
```



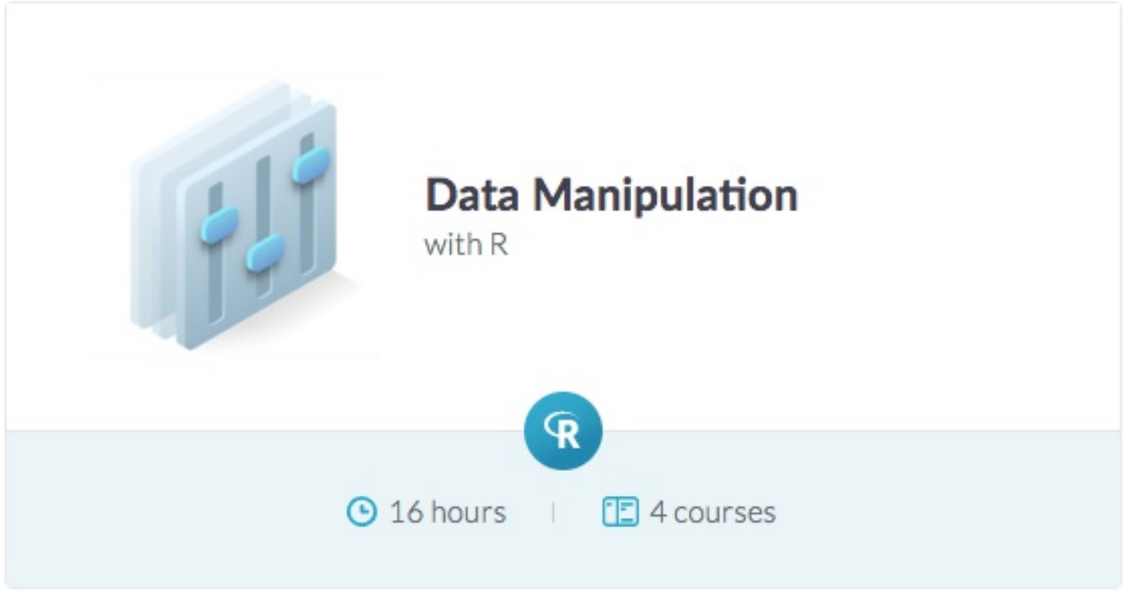
Next steps: Data visualization

- Data visualization with ggplot2



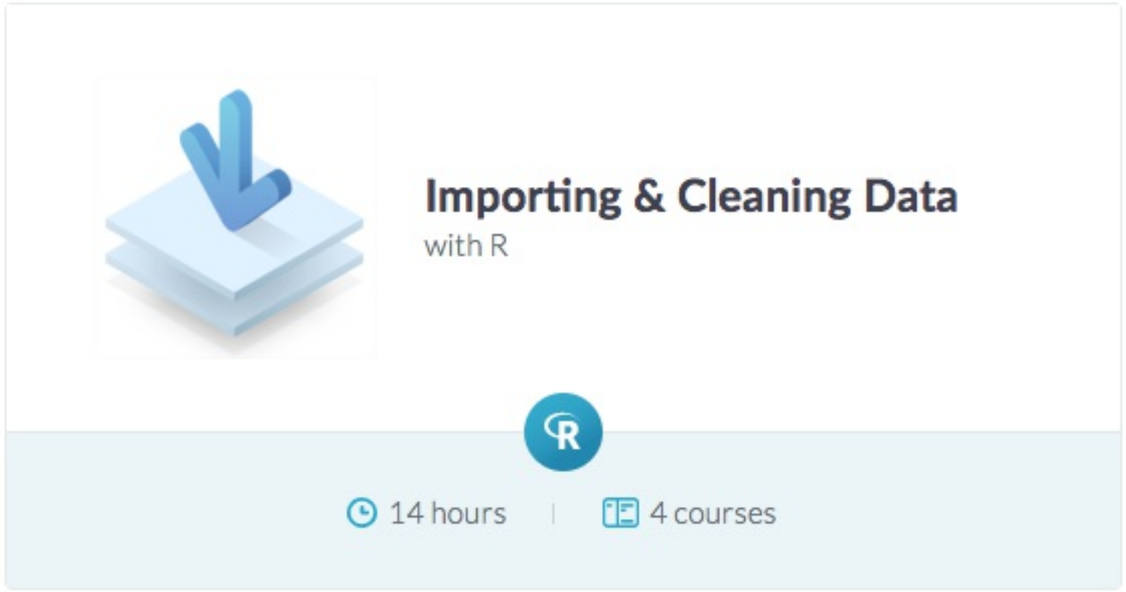
Next steps: Data manipulation

- Data manipulation with dplyr



Next steps: Importing and cleaning data

- Importing and cleaning data



Next steps: Practice!

- Exploratory Data Analysis in R: Case Study

INTRODUCTION TO THE TIDYVERSE

Enjoy your data science journey!