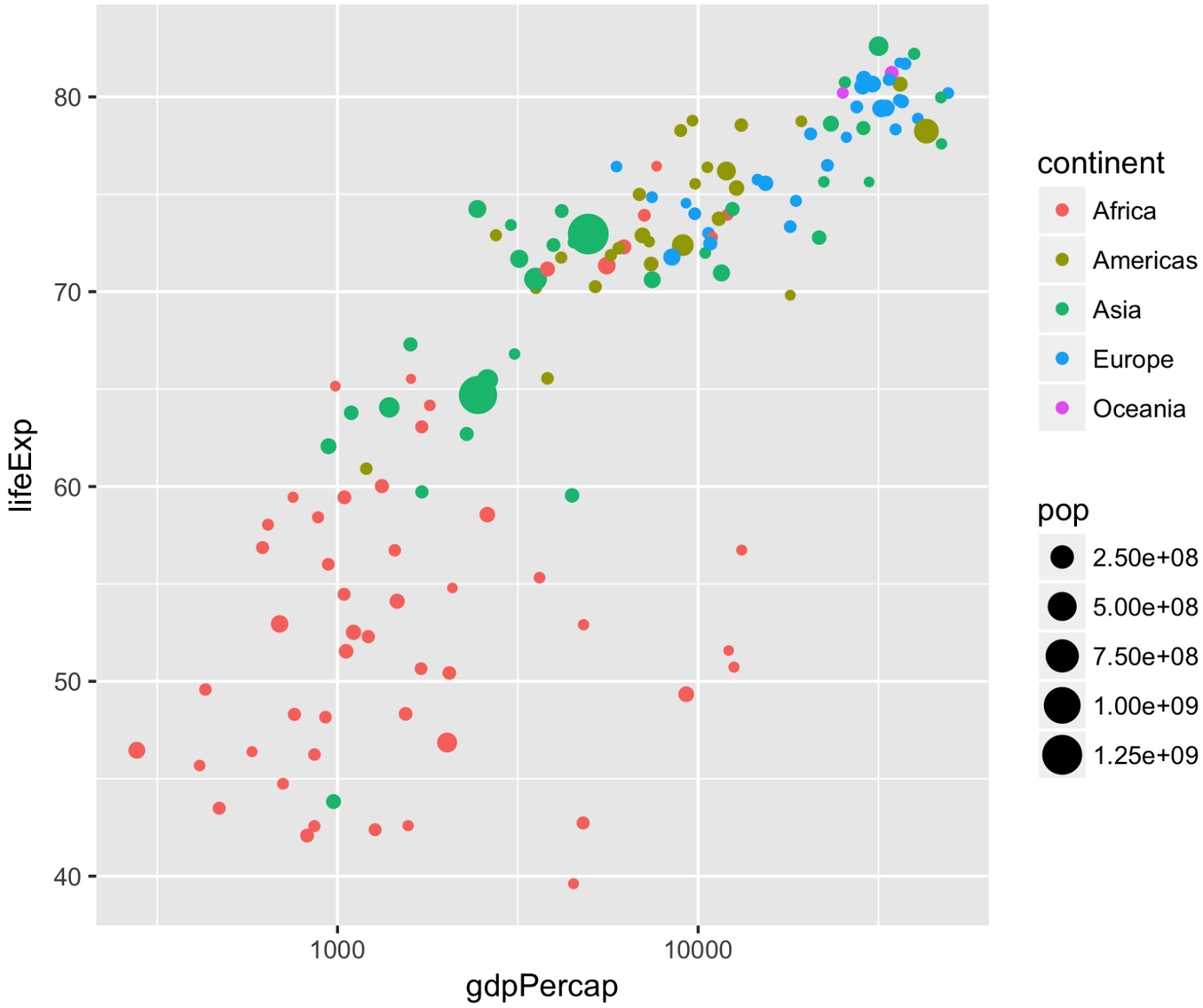
Visualizing with ggplot2

INTRODUCTION TO THE TIDYVERSE

David Robinson
Chief Data Scientist, DataCamp

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

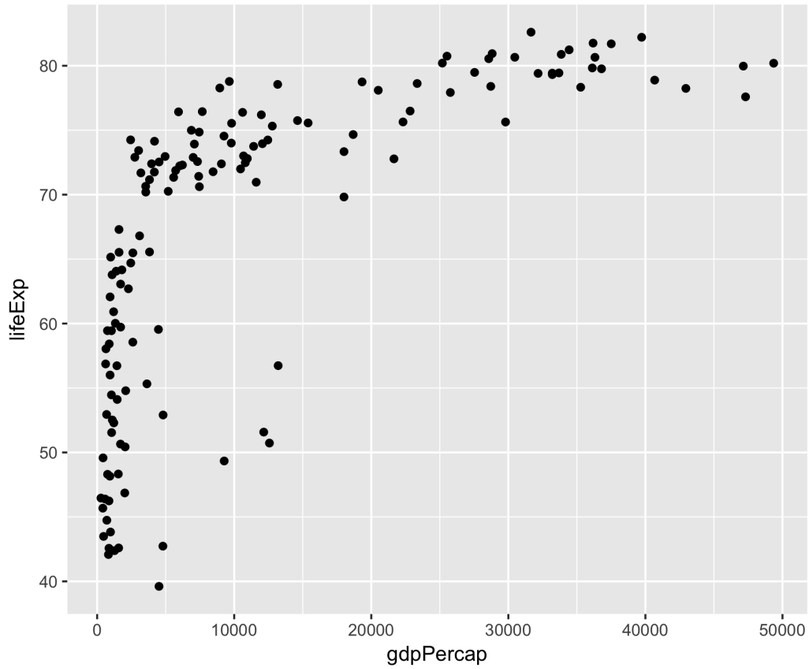
Variable Assignment

```
gapminder_2007 <- gapminder %>%
  filter(year == 2007)

gapminder_2007
```

```
# A tibble: 142 x 6
       country continent  year lifeExp       pop  gdpPercap
        <fctr>    <fctr> <int>   <dbl>     <int>      <dbl>
 1 Afghanistan      Asia  2007  43.828  31889923   974.5803
 2     Albania    Europe  2007  76.423   3600523  5937.0295
 3     Algeria    Africa  2007  72.301  33333216  6223.3675
 4      Angola    Africa  2007  42.731  12420476  4797.2313
 5   Argentina  Americas  2007  75.320  40301927 12779.3796
 6   Australia   Oceania  2007  81.235  20434176 34435.3674
 7     Austria    Europe  2007  79.829   8199783 36126.4927
 8     Bahrain      Asia  2007  75.635    708573 29796.0483
 9  Bangladesh      Asia  2007  64.062 150448339  1391.2538
10     Belgium    Europe  2007  79.441  10392226 33692.6051
# ... with 132 more rows
```

Visualizing with ggplot2




```
library(ggplot2)
```

```
ggplot(gapminder_2007, aes(x = gdpPerCap, y = lifeExp)) +
  geom_point()
```

Let's practice!

INTRODUCTION TO THE TIDYVERSE

DataCamp

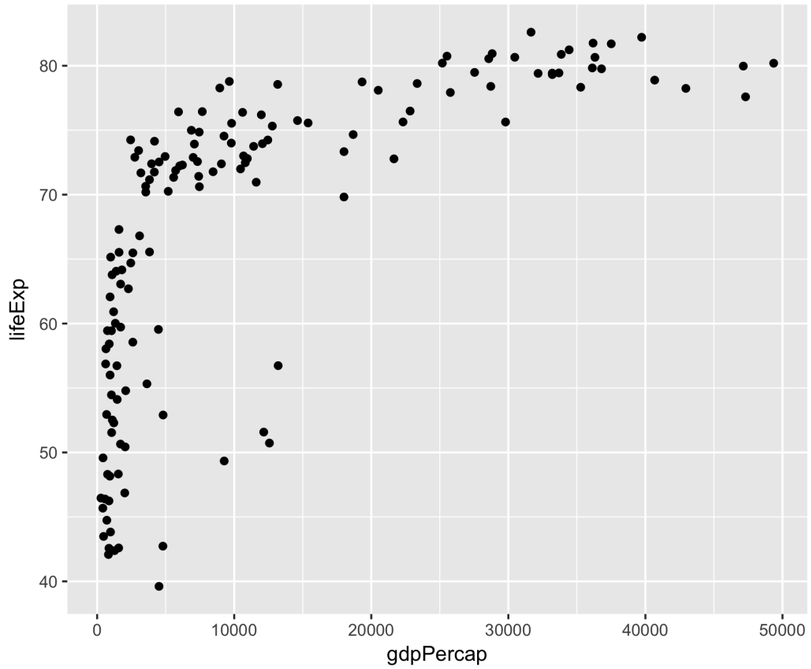
Log scales

INTRODUCTION TO THE TIDYVERSE

David Robinson
Chief Data Scientist, DataCamp

DataCamp

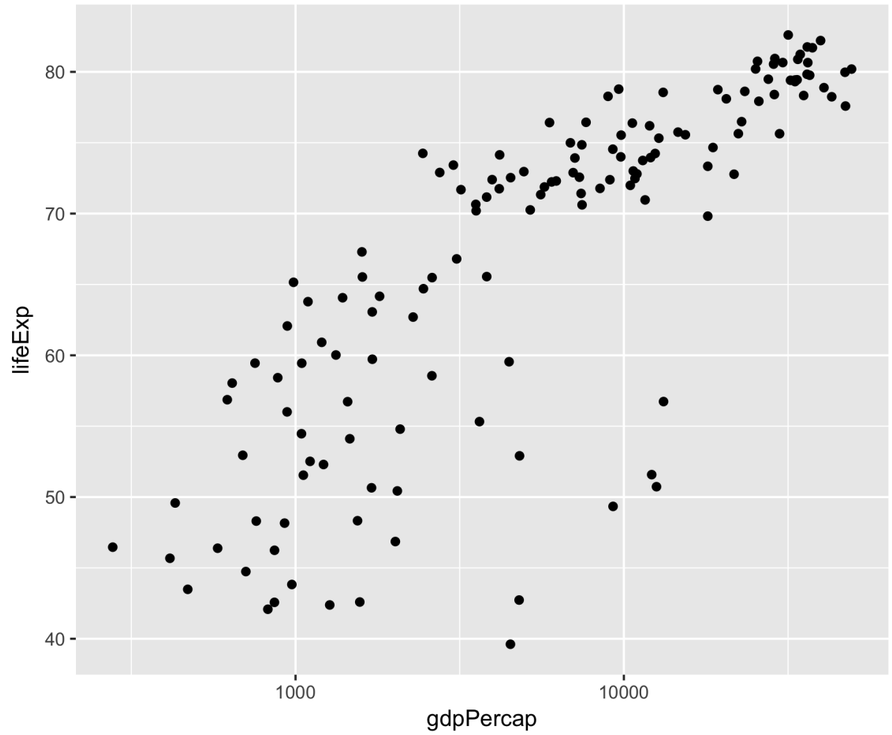
Scatter plot




```
library(ggplot2)

ggplot(gapminder, aes(x = gdpPerCap, y = lifeExp)) +
  geom_point()
```

Log scale




```
ggplot(gapminder_2007, aes(x = gdpPercap, y = lifeExp)) +
  geom_point() +
  scale_x_log10()
```

Let's practice!

INTRODUCTION TO THE TIDYVERSE

DataCamp

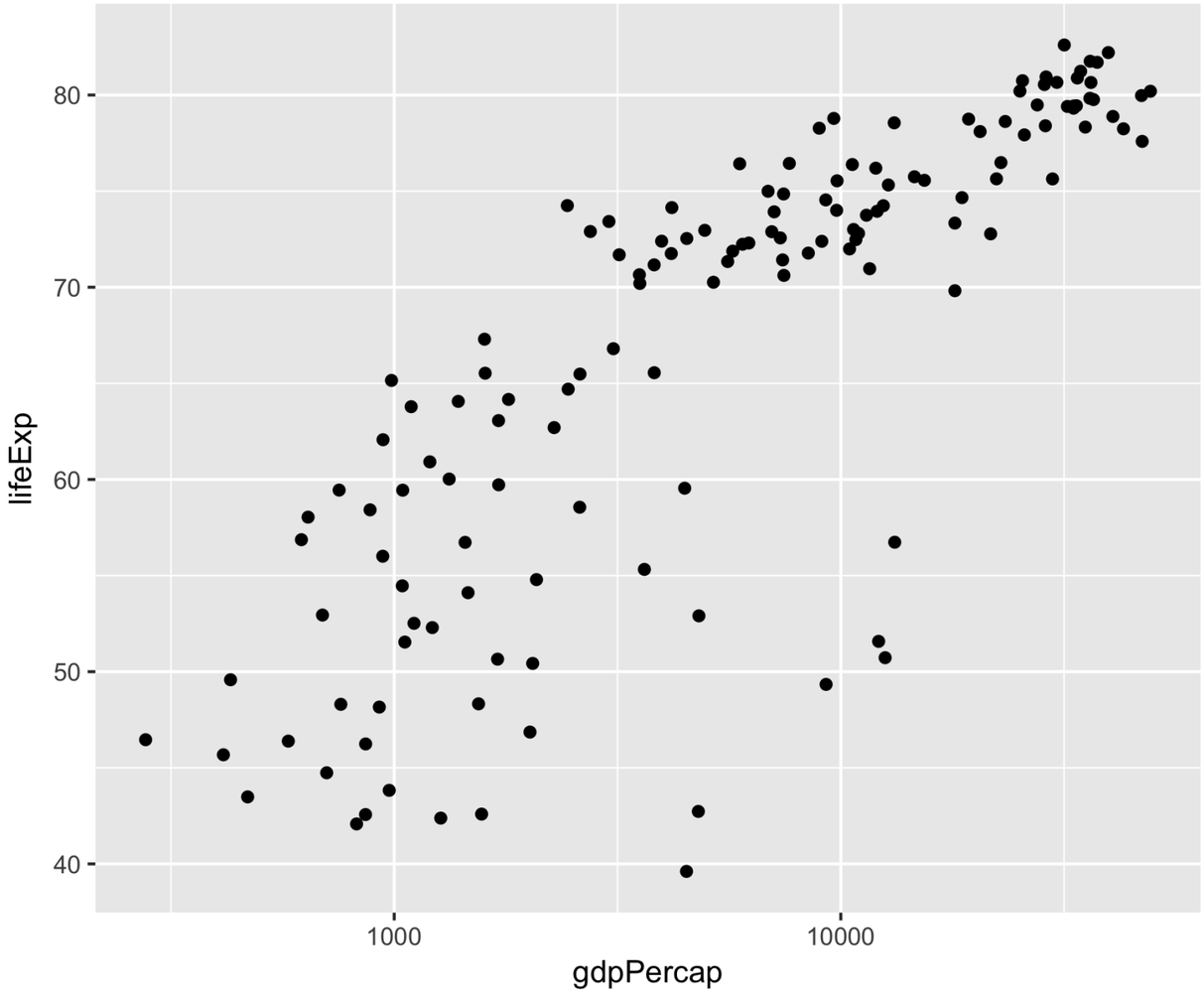
Additional aesthetics

INTRODUCTION TO THE TIDYVERSE

David Robinson
Chief Data Scientist, DataCamp

DataCamp

Scatter plots

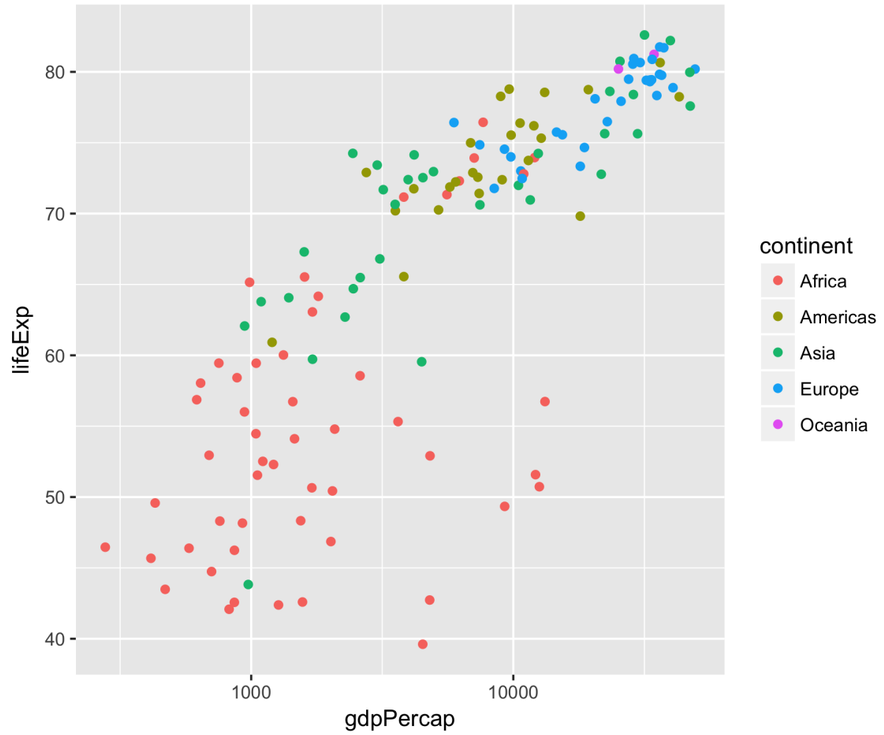


Additional variables

```
gapminder_2007
```

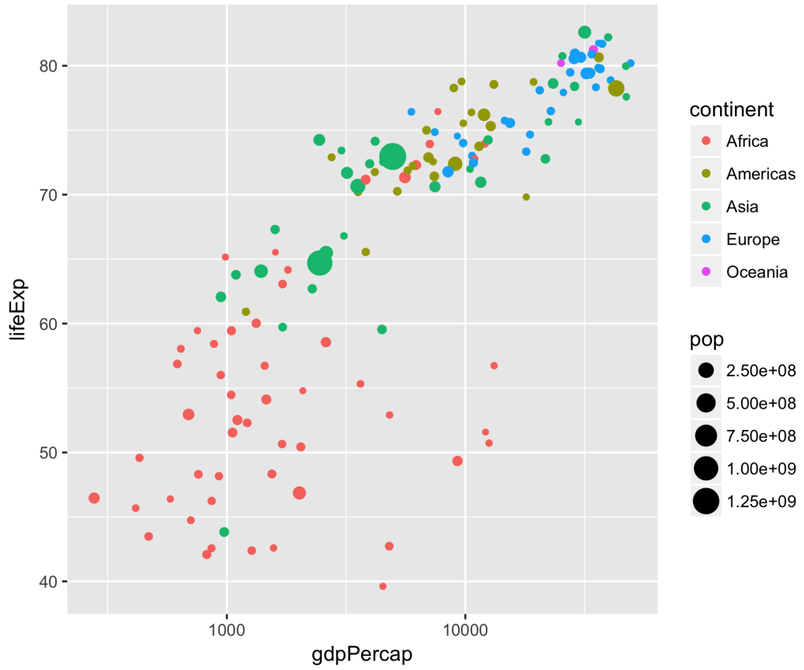
```
# A tibble: 142 x 6
       country continent  year lifeExp       pop  gdpPercap
        <fctr>    <fctr> <int>   <dbl>     <dbl>      <dbl>
 1 Afghanistan      Asia  2007  43.828  31889923   974.5803
 2     Albania    Europe  2007  76.423   3600523  5937.0295
 3     Algeria    Africa  2007  72.301  33333216  6223.3675
 4      Angola    Africa  2007  42.731  12420476  4797.2313
 5   Argentina  Americas  2007  75.320  40301927 12779.3796
 6   Australia   Oceania  2007  81.235  20434176 34435.3674
 7     Austria    Europe  2007  79.829   8199783 36126.4927
 8     Bahrain      Asia  2007  75.635    708573 29796.0483
 9  Bangladesh      Asia  2007  64.062 150448339  1391.2538
10     Belgium    Europe  2007  79.441  10392226 33692.6051
# ... with 132 more rows
```

The color aesthetic




```
ggplot(gapminder_2007, aes(x = gdpPercap, y = lifeExp, color = continent)) +
  geom_point() +
  scale_x_log10()
```

The size aesthetic




```
ggplot(gapminder_2007, aes(x = gdpPercap, y = lifeExp, color = continent,
                           size = pop)) +
  geom_point() +
  scale_x_log10()
```

Aesthetics

Aesthetic	Variable
x	gdpPerCap
y	lifeExp
color	continent
size	pop

Let's practice!

INTRODUCTION TO THE TIDYVERSE

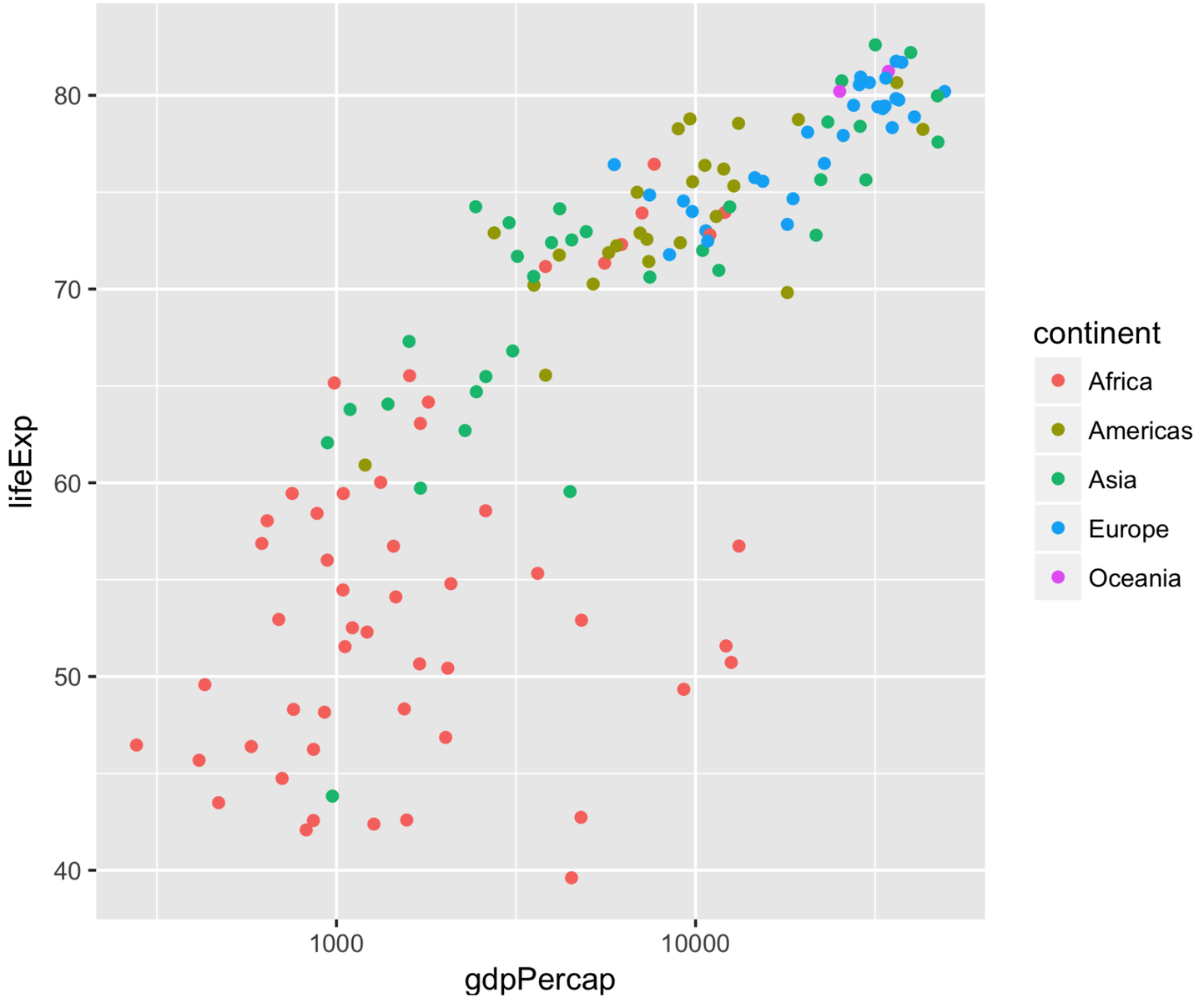
DataCamp

Faceting

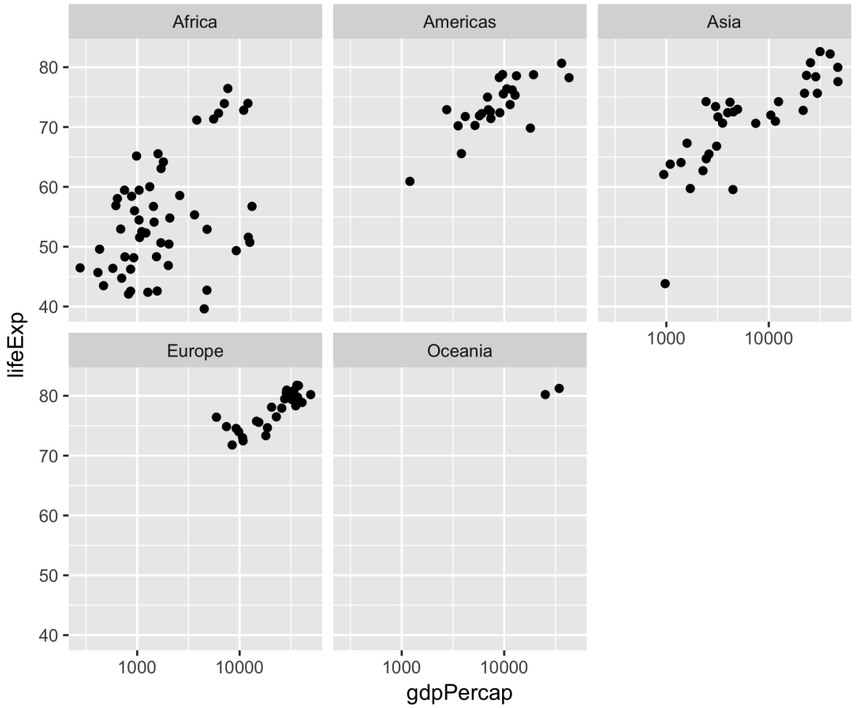
INTRODUCTION TO THE TIDYVERSE

David Robinson
Chief Data Scientist, DataCamp

DataCamp


80
70
60
50
40
lifeExp
1000
10000
gdpPercap
continent
Africa
Americas
Asia
Europe
Oceania

Faceting




```
ggplot(gapminder_2007, aes(x = gdpPercap, y = lifeExp)) +
  geom_point() +
  scale_x_log10() +
  facet_wrap(~ continent)
```

1952
1957
1962
1967
1972
1977
1982
1987
1992
1997
2002
2007
lifeExp
gdpPercap
80
60
40
1e+03
1e+04
1e+05
continent
Africa
Americas
Asia
Europe
Oceania
pop
2.50e+08
5.00e+08
7.50e+08
1.00e+09
1.25e+09

Let's practice!

INTRODUCTION TO THE TIDYVERSE

DataCamp