

# DATA VISUALIZATION WITH R

---

J. Alexander Branham

Fall 2017

# DATA VISUALIZATION

---

- Plots are (usually) easier to interpret than tables and numbers

- Plots are (usually) easier to interpret than tables and numbers
- We'll use **ggplot2** to visualize data in R. Make sure you have it installed (only the first time) and loaded (every time):

- Plots are (usually) easier to interpret than tables and numbers
- We'll use **ggplot2** to visualize data in R. Make sure you have it installed (only the first time) and loaded (every time):

- Plots are (usually) easier to interpret than tables and numbers
- We'll use **ggplot2** to visualize data in R. Make sure you have it installed (only the first time) and loaded (every time):

```
install.packages(c("ggplot2", "gapminder"))  
library(ggplot2, gapminder)
```

We'll be working with the `gapminder` dataset, provided by `gapminder`. Check out its documentation with `?gapminder`:

1. How many rows (observations) are in the data?

We'll be working with the `gapminder` dataset, provided by `gapminder`. Check out its documentation with `?gapminder`:

1. How many rows (observations) are in the data?
2. What's the name of the variable for life expectancy?



We'll be working with the `gapminder` dataset, provided by `gapminder`. Check out its documentation with `?gapminder`:

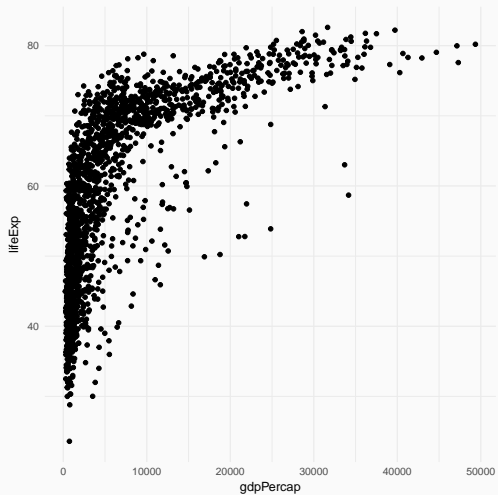
1. How many rows (observations) are in the data?
2. What's the name of the variable for life expectancy?
3. What's the name of the variable for GDP per capita?

Kuwait has a super high GDP per cap in some years. We're just going to drop them for the purpose of this tutorial:

```
gapminder <- gapminder %>%  
  filter(gdpPercap < 50000)
```

```
p <- ggplot(data = gapminder) +  
  geom_point(mapping = aes(x = gdpPercap, y = lifeExp))
```

# SCATTERPLOTS



# AESTHETICS

---

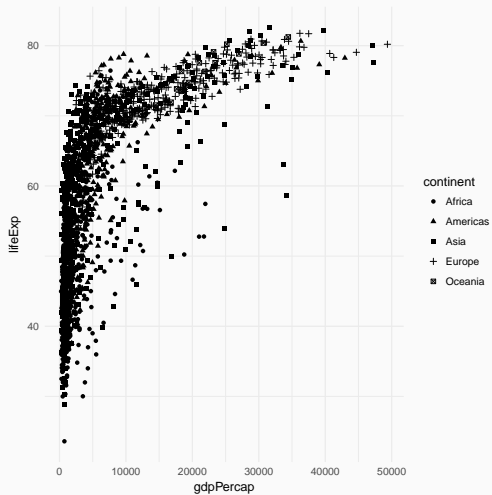
```
p <- ggplot(data = gapminder) +  
  geom_point(mapping = aes(x = gdpPercap,  
                           y = lifeExp,  
                           color = continent))
```

# COLOR



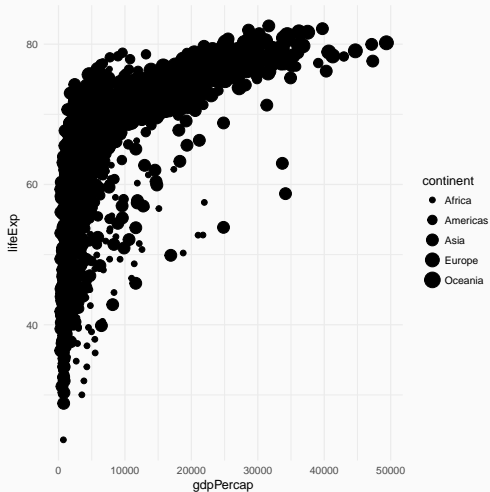
```
p <- ggplot(data = gapminder) +  
  geom_point(mapping = aes(x = gdpPercap,  
                           y = lifeExp,  
                           shape = continent))
```



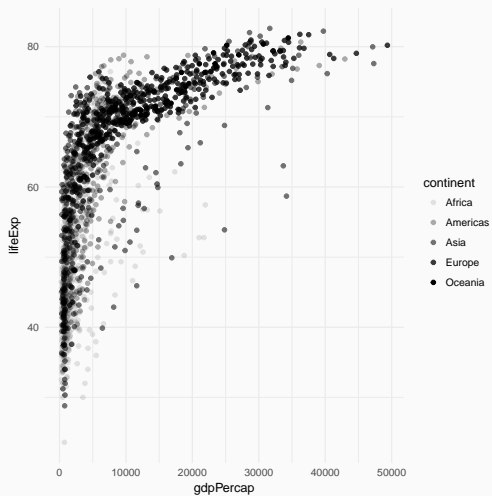


```
p <- ggplot(data = gapminder) +  
  geom_point(mapping = aes(x = gdpPercap,  
                           y = lifeExp,  
                           size = continent))
```

```
## Warning: Using size for a discrete variable is not advised.
```



```
p <- ggplot(data = gapminder) +  
  geom_point(mapping = aes(x = gdpPercap,  
                           y = lifeExp,  
                           alpha= continent))
```



## YOU TRY!

Use the `mpg` dataset from `ggplot2` to:

1. Make a scatterplot with `displ` on the x-axis and `hwy` on the y-axis

## YOU TRY!

Use the `mpg` dataset from `ggplot2` to:

1. Make a scatterplot with `displ` on the x-axis and `hwy` on the y-axis
2. Add color, size, and shape aesthetics

## FACETS

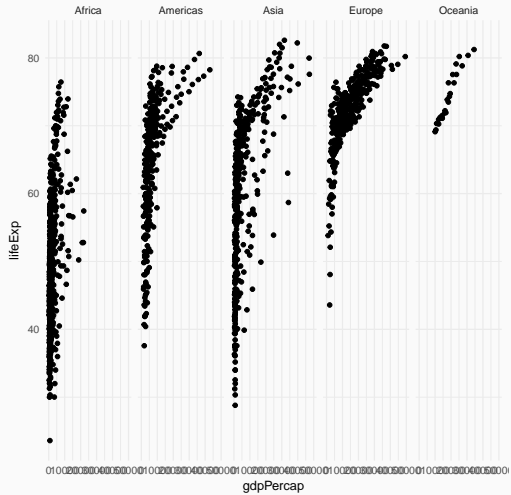
---



Can make facets by adding `facet_grid`:

```
p <- ggplot(data = gapminder) +  
  geom_point(mapping = aes(x = gdpPercap, y = lifeExp)) +  
  facet_grid(. ~ continent)
```

# FACETS



## GEOMS

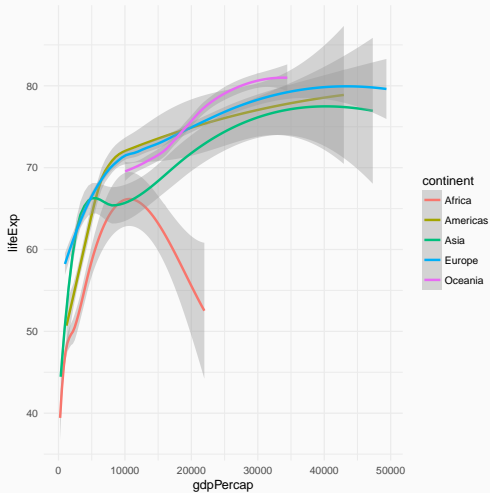
---

There are other kinds of graphs than scatterplots. `geom_` takes care of this for `ggplot2` (“geometric object”):

```
p <- ggplot(data = gapminder,  
            mapping = aes(x = gdpPercap, y = lifeExp, color = continent))  
      geom_smooth()
```

# SMOOTH

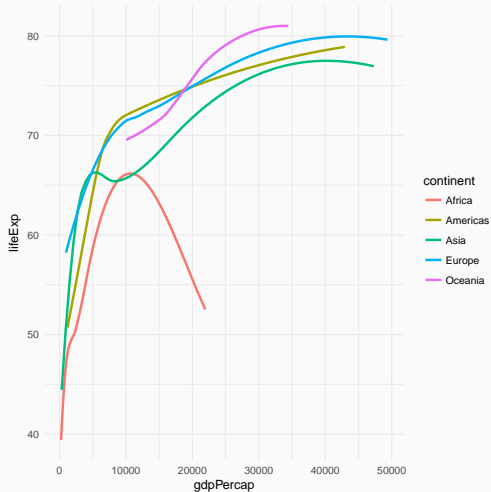
```
## `geom_smooth()` using method = 'loess'
```



```
p <- ggplot(data = gapminder,  
            mapping = aes(x = gdpPercap, y = lifeExp,  
                           color = continent)) +  
  geom_smooth(se = FALSE)
```

## SMOOTH - SEs

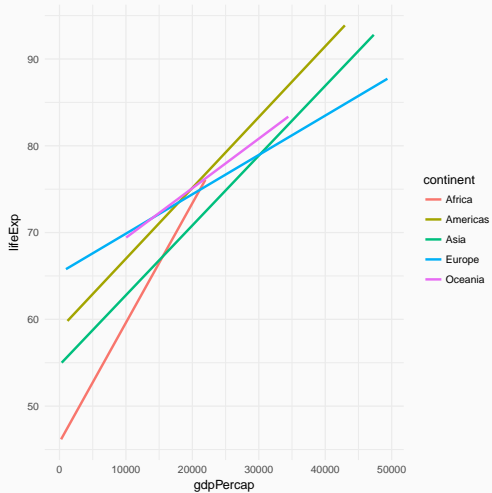
```
## `geom_smooth()` using method = 'loess'
```





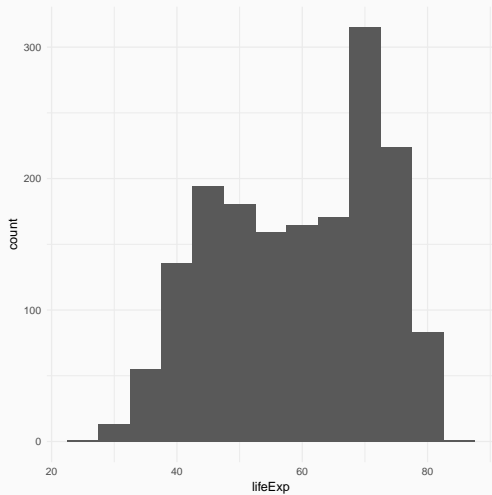
```
p <- ggplot(data = gapminder,  
            mapping = aes(x = gdpPercap, y = lifeExp,  
                          color = continent)) +  
  geom_smooth(se = FALSE, method = "lm")
```

# SMOOTH - OLS



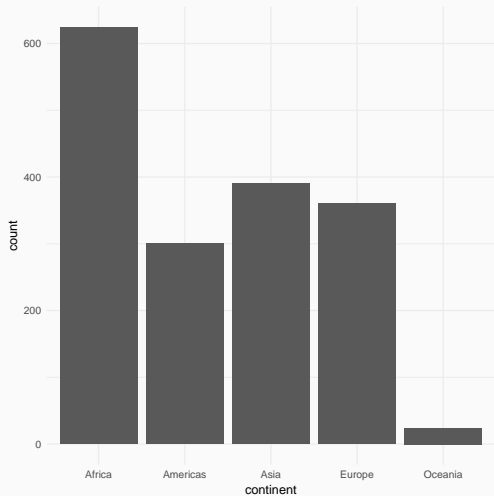
```
p <- ggplot(data = gapminder, mapping = aes(x = lifeExp)) +  
  geom_histogram(binwidth = 5)
```

# HISTOGRAM



```
p <- ggplot(data = gapminder, mapping = aes(x = continent)) +  
  geom_bar()
```

# BAR CHARTS

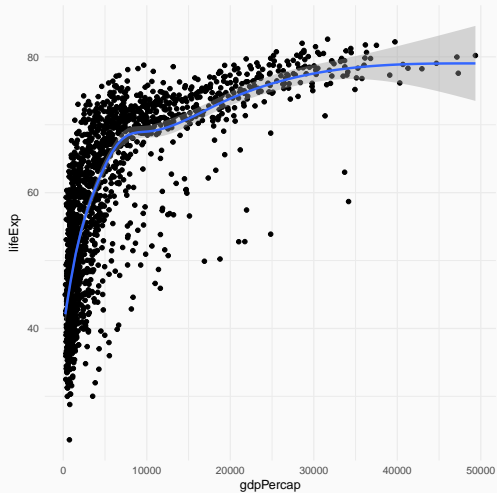


We can layer multiple geoms on top of each other:

```
p <- ggplot(data = gapminder, mapping = aes(x = gdpPercap, y = lifeExp))  
  geom_point() + geom_smooth()
```

## COMBINING GEOMS

```
## `geom_smooth()` using method = 'gam'
```





## YOU TRY!

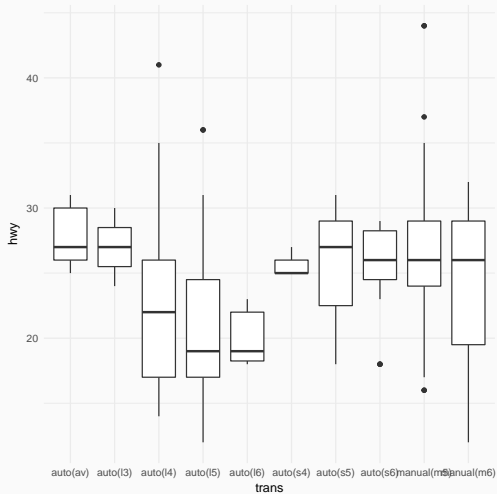
Using the `mpg` dataset, make a boxplot with the kind of transmission on the x-axis and the highway mpg on the y

## YOU TRY (ANSWERS)

```
p <- ggplot(data = mpg, mapping = aes(x = trans, y = hwy)) +  
  geom_boxplot()
```

## YOU TRY (ANSWERS)

p



## BAR CHARTS (EXTENDED)

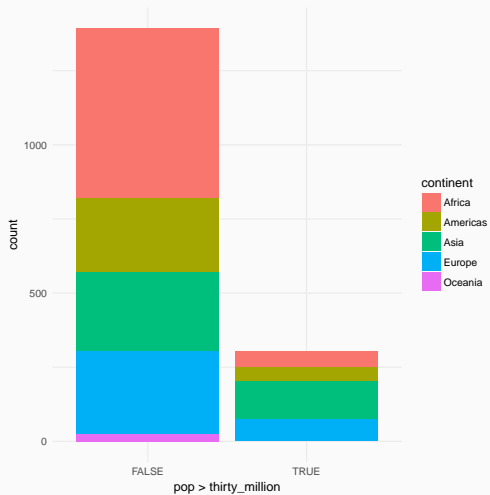
Fill vs color

What do you expect this to do:

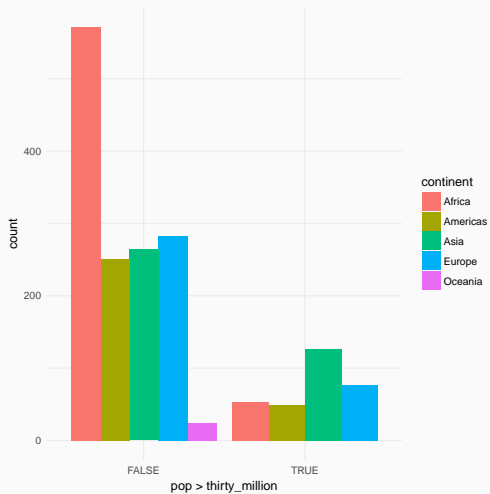
```
thirty_million <- 300000000  
ggplot(data = gapminder, mapping = aes(x = pop > thirty_million, color =  
  geom_bar()
```

## USE FILL FOR BAR CHARTS

```
p <- ggplot(data = gapminder, mapping = aes(x = pop > thirty_million,  
                                             fill = continent)) +  
  geom_bar()
```



Try to make this plot



```
ggplot(data = gapminder, mapping = aes(x = pop > thirty_million,  
                                         fill = continent)) +  
  geom_bar(position = "dodge")
```

There's "stack", "dodge", "identity", "fill" - try them all out!