Data Visualization
Kieran Healy, Ph.D.

Upcoming Seminar:
December 1-2, 2017, Philadelphia, Pennsylvania
Sample Slides

DATA VISUALIZATION

Kieran Healy
Using ggplot
HOW IT WORKS
<table>
<thead>
<tr>
<th>gdp</th>
<th>lifexp</th>
<th>pop</th>
<th>continent</th>
</tr>
</thead>
<tbody>
<tr>
<td>340</td>
<td>65</td>
<td>31</td>
<td>Euro</td>
</tr>
<tr>
<td>227</td>
<td>51</td>
<td>200</td>
<td>Amer</td>
</tr>
<tr>
<td>909</td>
<td>81</td>
<td>80</td>
<td>Euro</td>
</tr>
<tr>
<td>126</td>
<td>40</td>
<td>20</td>
<td>Asia</td>
</tr>
</tbody>
</table>
A Gapminder Plot

Life Expectancy vs. log GDP

Continent:
- Asia
- Euro
- Amer

Population:
- 0-35
- 36-100
- >100
1. Tidy Data

<table>
<thead>
<tr>
<th>gdp</th>
<th>lifexp</th>
<th>pop</th>
<th>continent</th>
</tr>
</thead>
<tbody>
<tr>
<td>340</td>
<td>65</td>
<td>31</td>
<td>Euro</td>
</tr>
<tr>
<td>227</td>
<td>51</td>
<td>200</td>
<td>Amer</td>
</tr>
<tr>
<td>909</td>
<td>81</td>
<td>80</td>
<td>Euro</td>
</tr>
<tr>
<td>126</td>
<td>40</td>
<td>20</td>
<td>Asia</td>
</tr>
</tbody>
</table>

x=gdp  y=lifexp  size=pop  color=continent

2. Mapping

```
ggplot(mapping = aes(x = ...))
```

3. Geom

```
geom_point()
```
4. Coordinate System

5. Scales

6. Labels & Guides

A Gapminder Plot

Continent

Population

Life Expectancy

log GDP
LAYER BY LAYER
head(gapminder)

```r
## # A tibble: 6 × 6
## #  country continent year  lifeExp  pop    gdpPercap
## # <fct> <fct>   <int>    <dbl> <int>      <dbl>
## 1 Afghanistan Asia     1952     28.801 8425333  779.4453
## 2 Afghanistan Asia     1957     30.332 9240934  820.8530
## 3 Afghanistan Asia     1962     31.997 10267083 853.1007
## 4 Afghanistan Asia     1967     34.020 11537966 836.1971
## 5 Afghanistan Asia     1972     36.088 13079460 739.9811
## 6 Afghanistan Asia     1977     38.438 14880372 786.1134
```

dim(gapminder)

```r
## [1] 1704  6
```
p <- ggplot(data = gapminder)
Mapping: tell ggplot the relationships you want to see

```r
p <- ggplot(data = gapminder,
            mapping = aes(x = gdpPercap,
                           y = lifeExp))
```
• The mapping = aes(...) instruction links variables to things you will see on the plot.

• The x and y values are the most obvious ones.

• Other aesthetic mappings can include, e.g., color, shape, and size.

• These mappings are not directly specifying what specific, e.g., colors or shapes will be on the plot. Rather they say which variables in the data will be represented by, e.g., colors and shapes on the plot.
p + geom_point()
p + geom_smooth()
This process is literally additive
p + geom_point() + geom_smooth(method = "lm")

Every geom is a function. It can take arguments.
```r
p <- ggplot(data = gapminder,
           mapping = aes(x = gdpPercap,
                          y = lifeExp))

p + geom_point() +
   geom_smooth(method = 'gam') +
   scale_x_log10()
```
p + geom_point() +
  geom_smooth(method='gam') +
  scale_x_log10(labels = scales::dollar) +
  labs(x = "GDP Per Capita",
       y = "Life Expectancy in Years",
       title = "Economic Growth and Life Expectancy",
       subtitle = "Data points are country-years")
Economic Growth and Life Expectancy

Data points are country-years

GDP Per Capita

Life Expectancy in Years