GRD 610A Data Visualization II

Basic Charts in ggplot

Jenn Schilling

February 3, 2021

Today

- Visualization of the Week
- Discussion of Introduction and Chapter 1 of *The Truthful Art* (Cairo)
- Basic Charts in ggplot - Chapter 3 of *Data Visualization* (Healy)

Revisit Padlet

Data Visualization II Padlet

What should we add?

What new information have we gained this week?

Discussion

“A good visualization is:

1. reliable information,
2. visually encoded so relevant patterns become noticeable,
3. organized in a way that enables at least some exploration, when it’s appropriate,
4. and presented in an attractive manner, but always remembering that honesty, clarity, and depth come first.”

From p. 12 of Cairo, A. (2016). *The truthful art: Data, charts, and maps for communication*. New Riders.

Do you agree? Should anything be added?

Discussion

Candid communication: "begin with the information and then…thoroughly analyze it to discover the messages worth spreading" (2016, Ciaro, *The Truthful Art*, p. 15)

What do you think of the elements of journalism (p. 21)?

How do these elements relate to this class?

Discussion

How do you define data visualization versus infographic?

What aspects of the data visualizations and infographics presented in Chapter 1 of *The Truthful Art* stood out to you?

Some Interactive Examples

Periscope's "A World of Terror"

ProPublica's "Treatment Tracker"

Wall Street Journal's "HealthCare.gov Explorer"

The Guardian's "Beyond the Border"

15 Minute Break

15:00

Chapter 3: Make A Plot

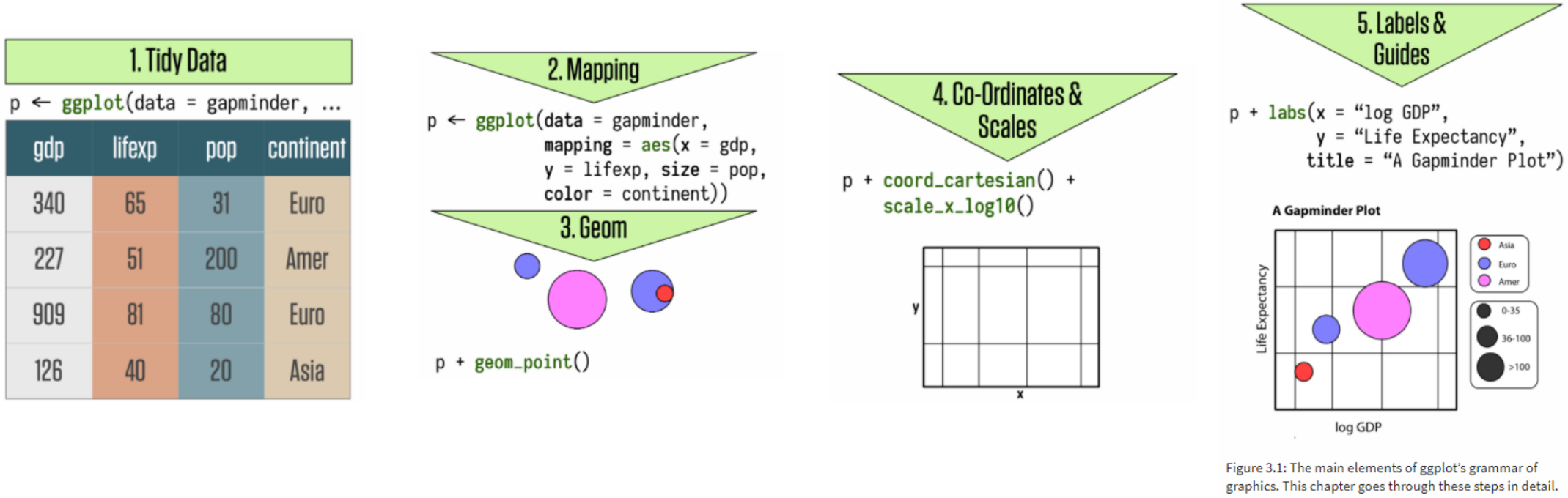


Figure 3.1: The main elements of ggplot's grammar of graphics. This chapter goes through these steps in detail.

From p. 56 of Healy, K. (2019). *Data visualization: A practical introduction*. Princeton ; Oxford: Princeton University Press.

Tidy Data

"In tidy data:

1. Each variable forms a column.
2. Each observation forms a row.
3. Each type of observational unit forms a table."

From p. 4 of Wickham, H. (2014). Tidy Data. *Journal of Statistical Software*, 59(10).

Wide Data

country	1952	1957	1962	1967	1972	1977	1982	1987	1992	1997	2002	2007
Afghanistan	28.801	30.332	31.997	34.020	36.088	38.438	39.854	40.822	41.674	41.763	42.129	43.828
Albania	55.230	59.280	64.820	66.220	67.690	68.930	70.420	72.000	71.581	72.950	75.651	76.423
Algeria	43.077	45.685	48.303	51.407	54.518	58.014	61.368	65.799	67.744	69.152	70.994	72.301
Angola	30.015	31.999	34.000	35.985	37.928	39.483	39.942	39.906	40.647	40.963	41.003	42.731
Argentina	62.485	64.399	65.142	65.634	67.065	68.481	69.942	70.774	71.868	73.275	74.340	75.320
Australia	69.120	70.330	70.930	71.100	71.930	73.490	74.740	76.320	77.560	78.830	80.370	81.235

Long Data

country	year	lifeExp
Afghanistan	1952	28.801
Afghanistan	1957	30.332
Afghanistan	1962	31.997
Afghanistan	1967	34.020
Afghanistan	1972	36.088
Afghanistan	1977	38.438

Tidy Data

country	continent	year	lifeExp	pop	gdpPercap
Afghanistan	Asia	1952	28.801	8425333	779.4453
Afghanistan	Asia	1957	30.332	9240934	820.8530
Afghanistan	Asia	1962	31.997	10267083	853.1007
Afghanistan	Asia	1967	34.020	11537966	836.1971
Afghanistan	Asia	1972	36.088	13079460	739.9811
Afghanistan	Asia	1977	38.438	14880372	786.1134

Mapping Data

```
p <- ggplot(data = gapminder)

p
```

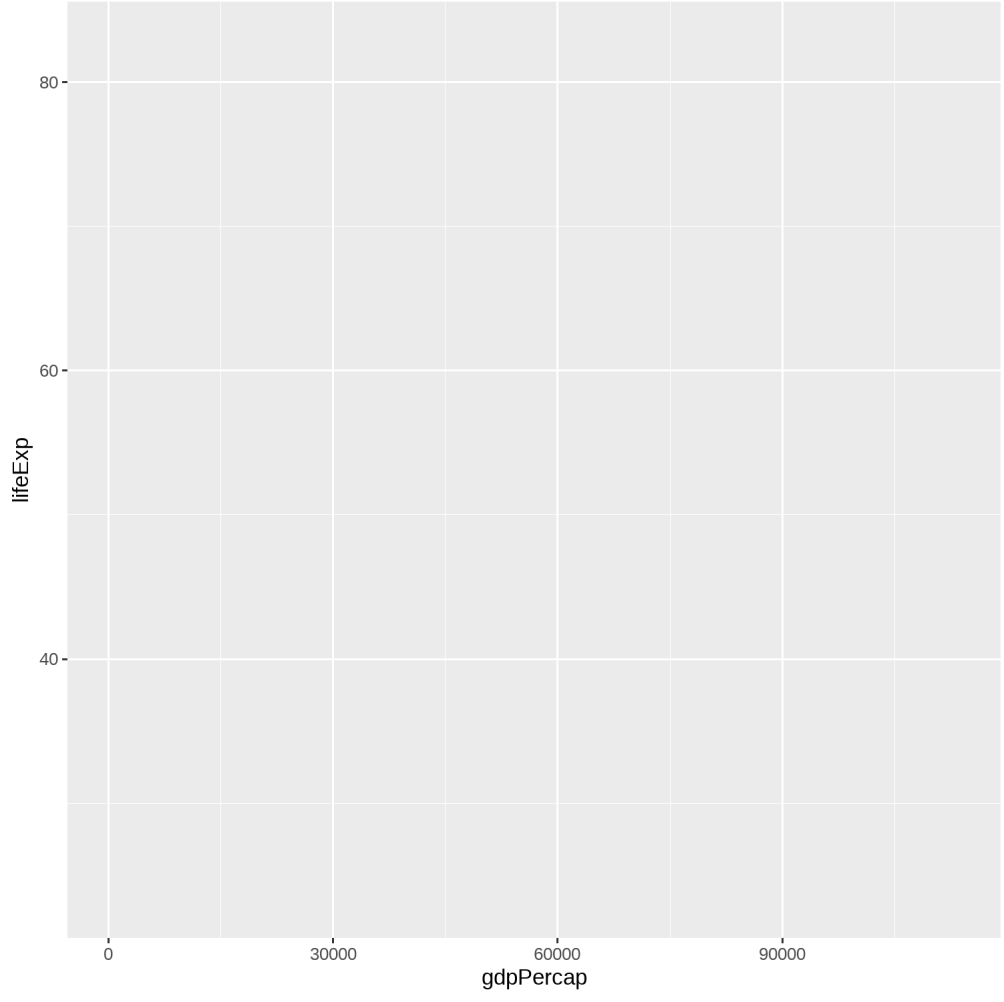
ggplot knows the data, but not how to map it.

Mapping Data

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

p
```

- `aes()` links variables to parts of the plot
- ggplot knows the data and how to map it
- BUT we have not provided information about what to draw

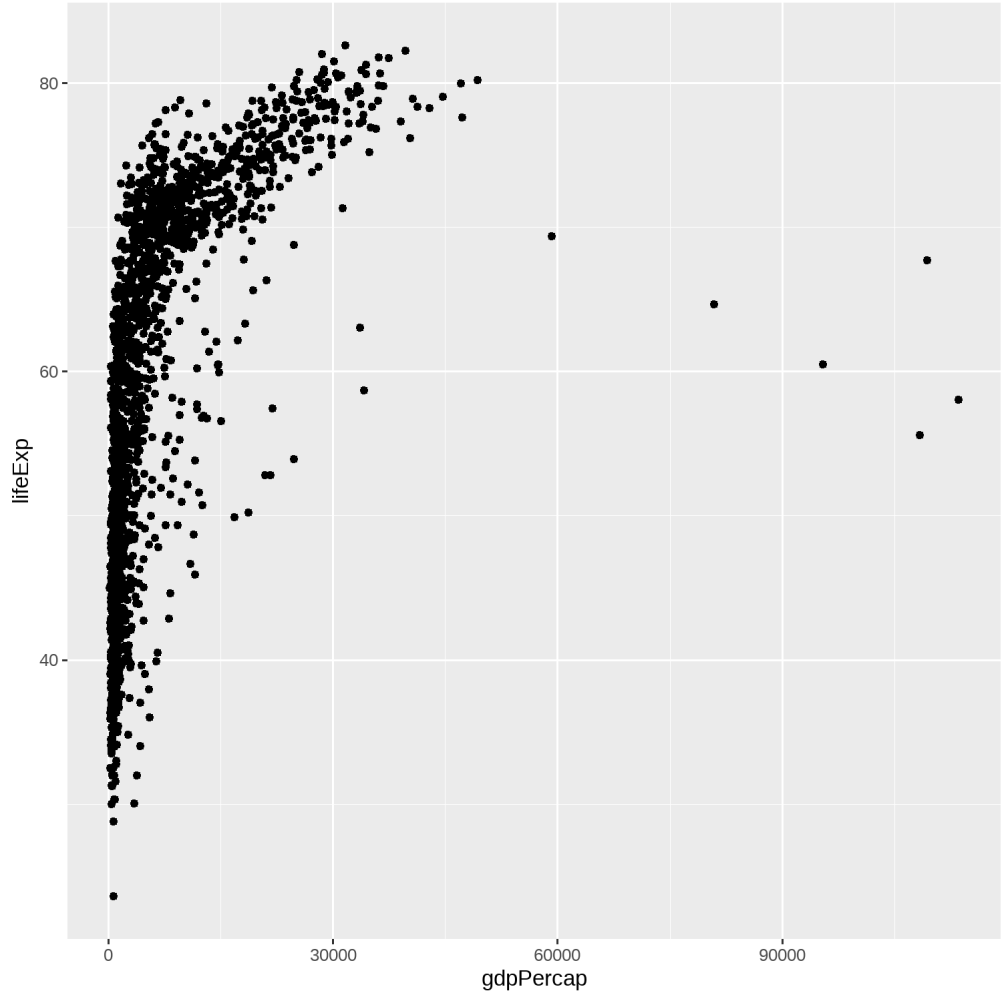


Drawing Data

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

p + geom_point()
```

Now ggplot knows the data, how to map it, and how to draw the plot (`geom_point()`)



The `ggplot` Process

1. Tell the `ggplot()` function what our data is. (`data = ...`)
2. Tell `ggplot()` what relationships we want to see.For convenience we will put the results of the first two steps in an object called `p`. (`mapping = aes(...)`)
3. Tell `ggplot` how we want to see the relationships in our data. (`geom_...`)
4. Layer on geoms as needed, by adding them to the `p` object one at a time.
5. Use some additional functions to adjust scales, labels, tick marks, titles. We'll learn more about some of these functions shortly.

From p. 60 of Healy, K. (2019). *Data visualization: A practical introduction*. Princeton ; Oxford: Princeton University Press.

Lab Time

Tasks to Complete

- Reading (see Syllabus)
- Prepare for your Visualization of the Week
- Practice R