

Intro to ggplot

Professor Halterman

Michigan State University

PLS 397 Analyzing and Visualizing Data
Fall 2023

- ▶ Sorry about the slow start! But I'm really happy with how prepared you are, so we'll be fine!
- ▶ Reading for Wednesday (covers what we learn today): sections 3.3, 3.4, 3.5
- ▶ We'll begin our short participation check-ins on Wednesday.

Let's make some plots!

`ggplot2` is an R library for making visualizations.

There's a logic behind how `ggplot` works—we're going to understand its logic by making some plots.

Loading the data

- ▶ Run the first part of the RMarkdown to load packages, including the `gapminder` library.
- ▶ You'll now have a dataframe called `gapminder`. We're just looking at one year of data for now.
- ▶ How many rows and columns does the data have? What are the variable names?

The building blocks of ggplot

```
ggplot(data = gapminder,  
       aes(x = gdpPercap, y = lifeExp))
```

The building blocks of ggplot

**ggplot is a function
that takes a few
arguments**



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

The building blocks of ggplot

The first argument is the data we're using.
(Note that you can leave out the `data=` part.)

```
ggplot(data = gapminder,  
       aes(x = gdpPercap, y = lifeExp))
```

The building blocks of ggplot

```
ggplot(data = gapminder,  
       aes(x = gdpPercap, y = lifeExp))
```



**Next, we define the “aesthetics” of the plot.
This is another function inside the main ggplot function.**

The building blocks of ggplot

```
ggplot(data = gapminder,  
       aes(x = gdpPercap, y = lifeExp))
```

The arguments to `aes` tell us which columns get mapped to which visual features.

The building blocks of ggplot


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

Another way to think about it:

- the x dimension will represent GDP
- the y dimension will represent life expectancy

The building blocks of ggplot

```
ggplot(data = gapminder,  
       aes(x = gdpPercap, y = lifeExp))
```

The `gdpPercap` column gets mapped to the x dimension

The building blocks of ggplot

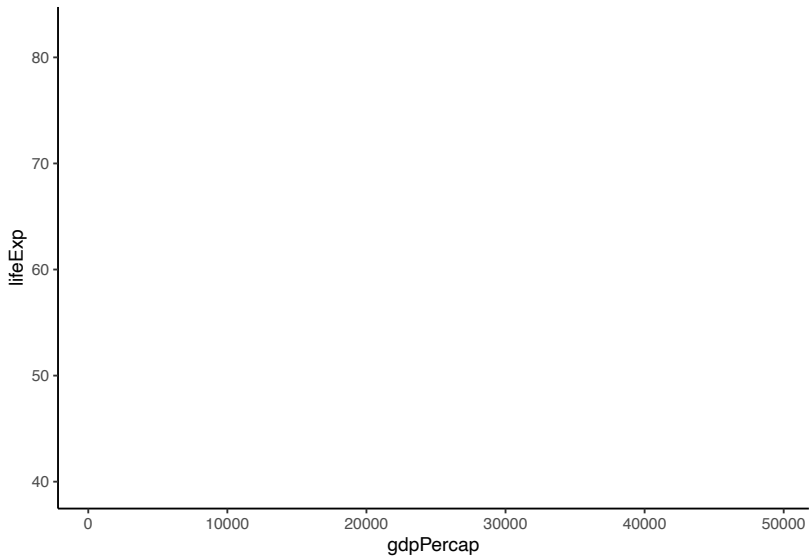
```
ggplot(data = gapminder,  
       aes(x = gdpPercap, y = lifeExp))
```

The `gdpPercap` column gets mapped to the x dimension

...and `lifeExp` gets mapped to y.

Now run that code!

Our first plot



Why no data?

We've told ggplot what data to use and how we want our variables mapped to visual aspects of the plot.

But we don't see anything? Why not?

We haven't told ggplot to draw anything yet!

To do that, we need to tell ggplot what we want it to draw. The library calls these **geoms**.

Adding geoms


ggplot is about adding layers to our plot

Here, the first line sets up our data and aesthetics

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

Adding geoms

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



We can tell ggplot to add a new layer with a + symbol.

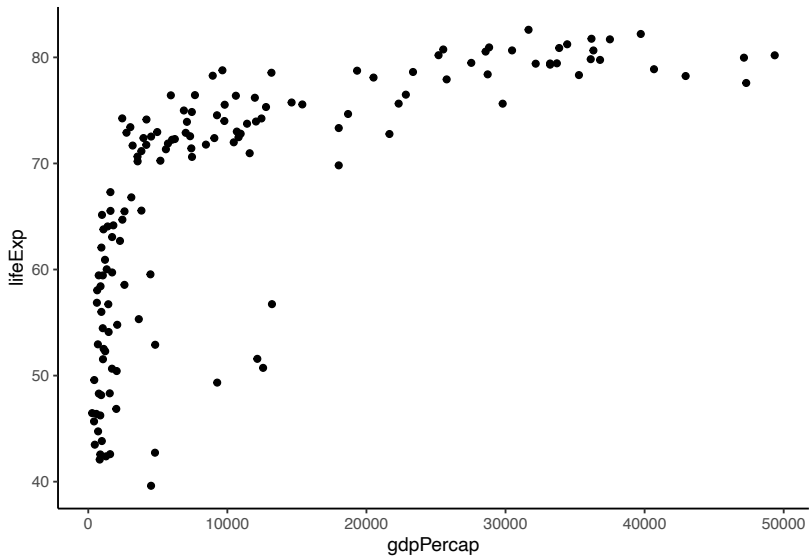
Adding geoms

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

We can tell it to use the x and y we just defined to draw points.

We do that with `geom_point`.

Our second plot



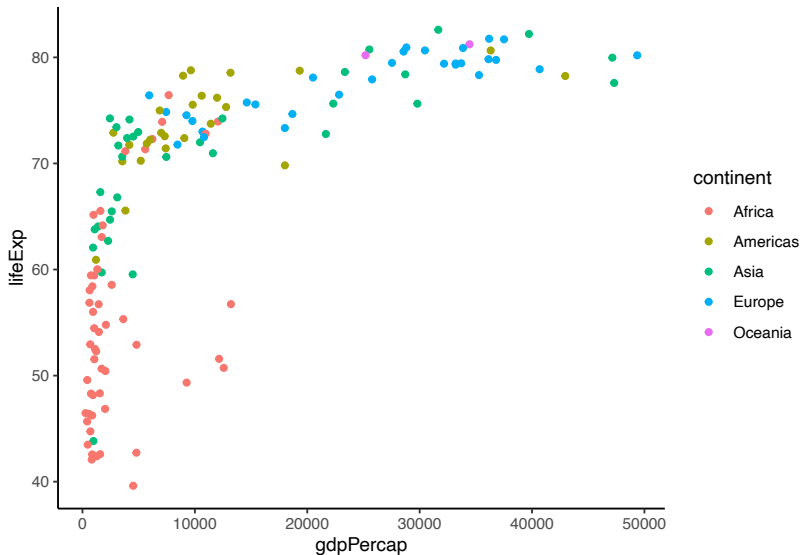
The philosophy of ggplot

- ▶ “**aesthetics**” map *columns* to *visual aspects*
 - “**What** we want to see”
 - Examples of other visual aspects?
- ▶ “**geoms**” display the data in a particular way
 - “**How** we want to see it”
 - We saw a *scatterplot*. Other examples?

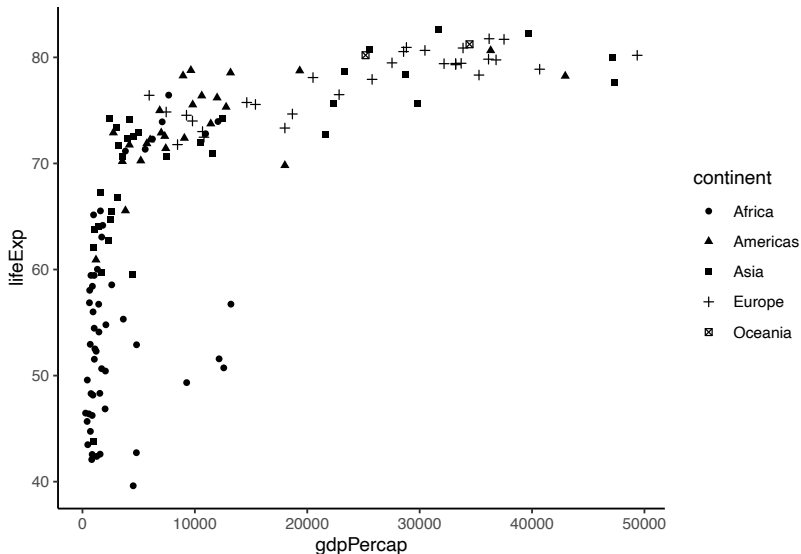
Other aesthetics

- ▶ We're exploring the relationship between GDP per capita and life expectancy.
- ▶ What if we were interested in understanding the role of **region** as well?
- ▶ Can we use a different aesthetic to show what part of the world a country is in?
- ▶ Hint: take a look at your data to find a column that represents this.

Showing continent with color



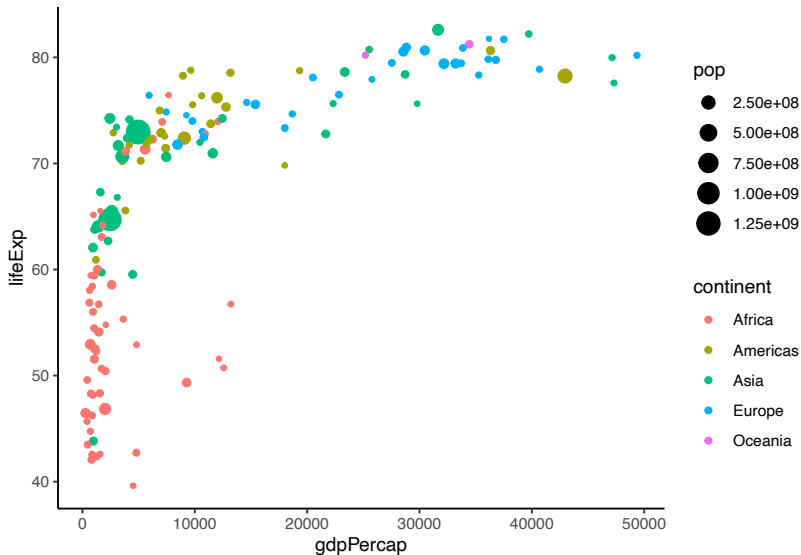
Showing continent with shape



Even more aesthetics

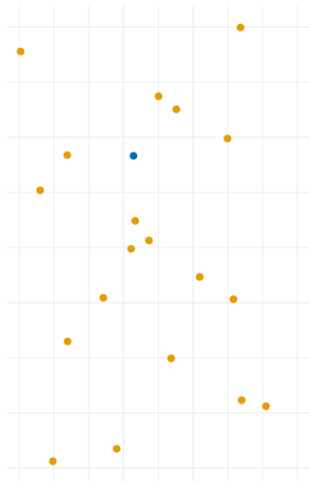
- ▶ Remember what we're doing when we make plots:
 - Exploring the data to draw insights
 - Thinking about how to communicate those insights to others
- ▶ Both of these require making judgements about what to show
- ▶ Example: do we want to show each country equally? Or do we want to emphasize larger countries?
- ▶ Think about how we could emphasize countries with larger populations

Showing population size

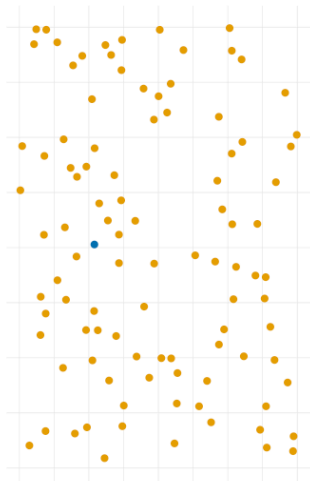


Visual decoding

Color Only, N=20



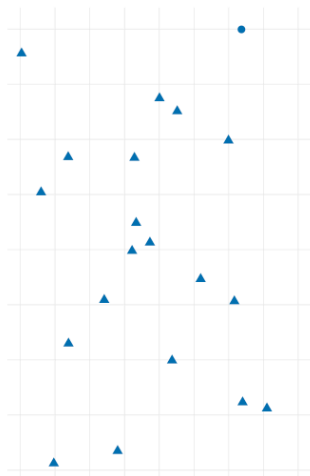
Color Only, N=100



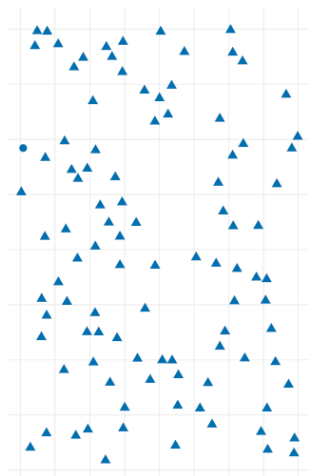
Credit: Healy, Data Visualization

Visual decoding

Shape Only, N=20



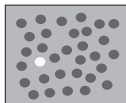
Shape Only, N=100



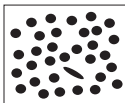
Data Visualization

Credit: Healy,

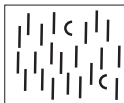
Making points pop



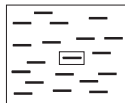
Grey value



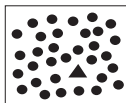
Elongation



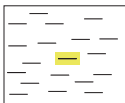
Curvature



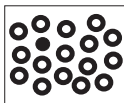
Added surround box



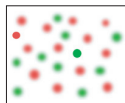
Shape



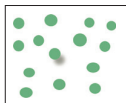
Added surround color



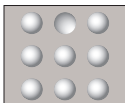
Filled



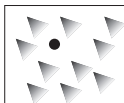
Sharpness



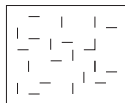
Cast shadow



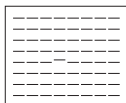
Convex and concave



Sharp vertex



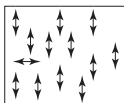
Joined lines



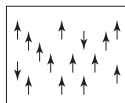
Misalignment



Blinking



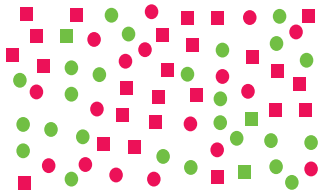
Direction of motion



Phase of motion

Credit: Ware, Visual Thinking for Design

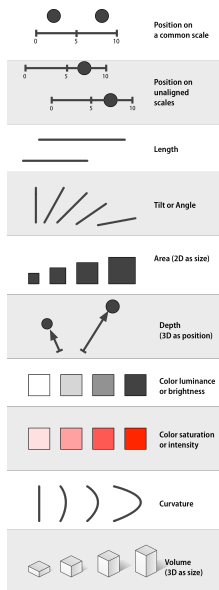
Visual decoding



There are three green squares in this pattern. The green squares do not show a pop-out effect, even though you know what to look for. The problem is that your primary visual cortex can either be tuned for the square shapes, or the green things, but not both.

Credit: Ware, Visual Thinking for Design

Research on visual decoding



Humans have an easier time with some aesthetics than others.

Cleveland and McGill (1984, 1987) and Munzer (2014) conducted experiments where they measured people's error in decoding data.

This figure shows the order of accuracy.

Summarizing aesthetics

- ▶ Aesthetics tell us what data we want to see on which visual “channel”
 - x, y
 - color
 - size
 - shape
 - fill (i.e, the center of a shape vs. its outline)
 - “alpha” (transparency)
 - linetype (dashed, solid, etc)
 - text (labels)
- ▶ Some aesthetics are more easily interpreted than others
- ▶ Less is often more: including shapes *and* color makes it very difficult to decode.
- ▶ Next time: more geoms!