# Fundamentals of Data Visualization

K Arnold, based on IntroDS.org

# Problem of Diverse Audiences

- Some of you have seen R, visualization, etc. before.

  > "This is still just review" -- says several people

- Some of you had CS 108 years ago...

  > "I'm totally lost" -- many people think, few say

-> **Yejae office hour today 4:30-5:30p**
-> My office hours **Fridays 3-4pm, Mondays 8-9am**.

# Q & A

> *When to use Console vs Rmarkdown?*

- **RMarkdown** is your source of truth. Use Console just to explore.
- Something broke? **Restart R and Run All**

> *Why two assignment operators?*

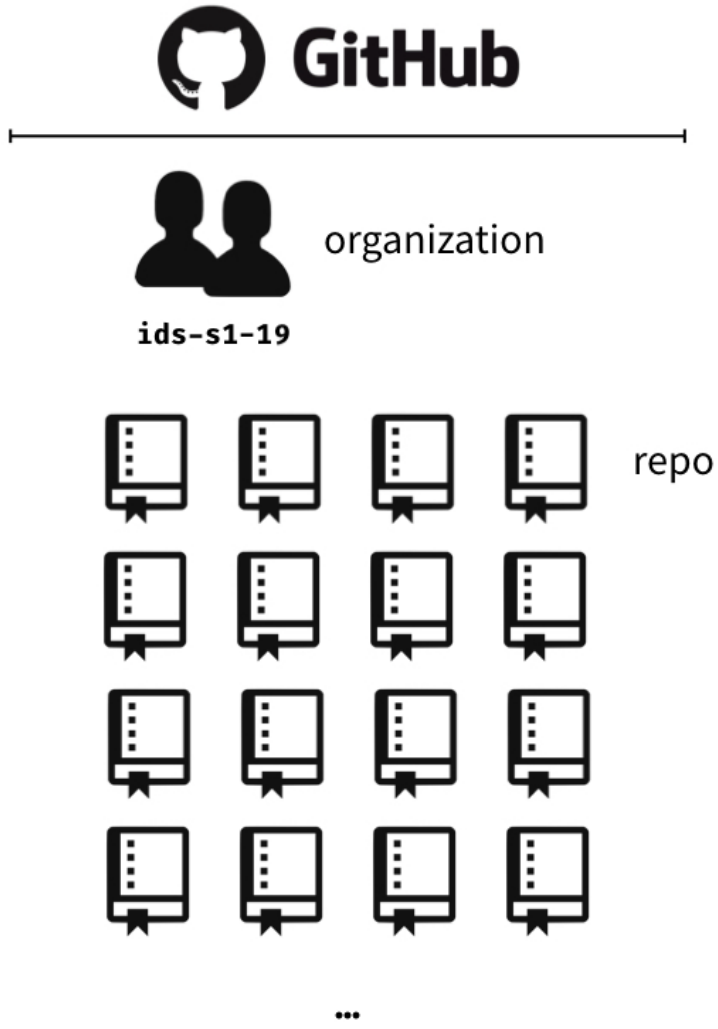
- `<-` assigns variables in *current environment*: `data <- read_csv()`
- `=` labels arguments to *functions*: `ggplot(data = dino_data)`
- You can use `=` instead of `<-`, but we won't.

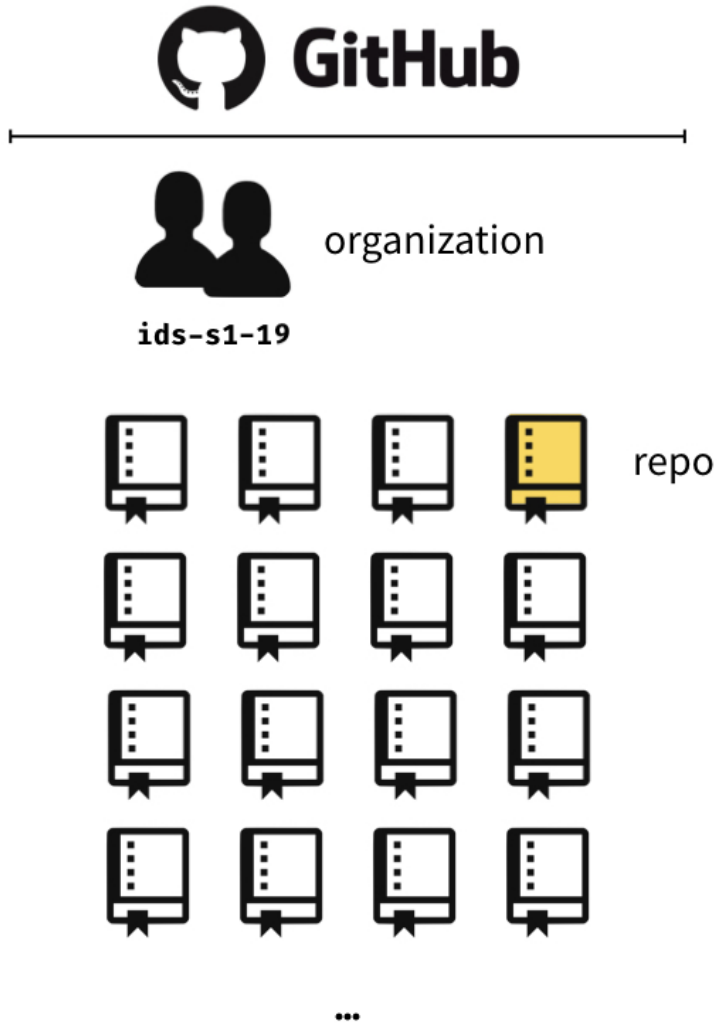
# Q & A

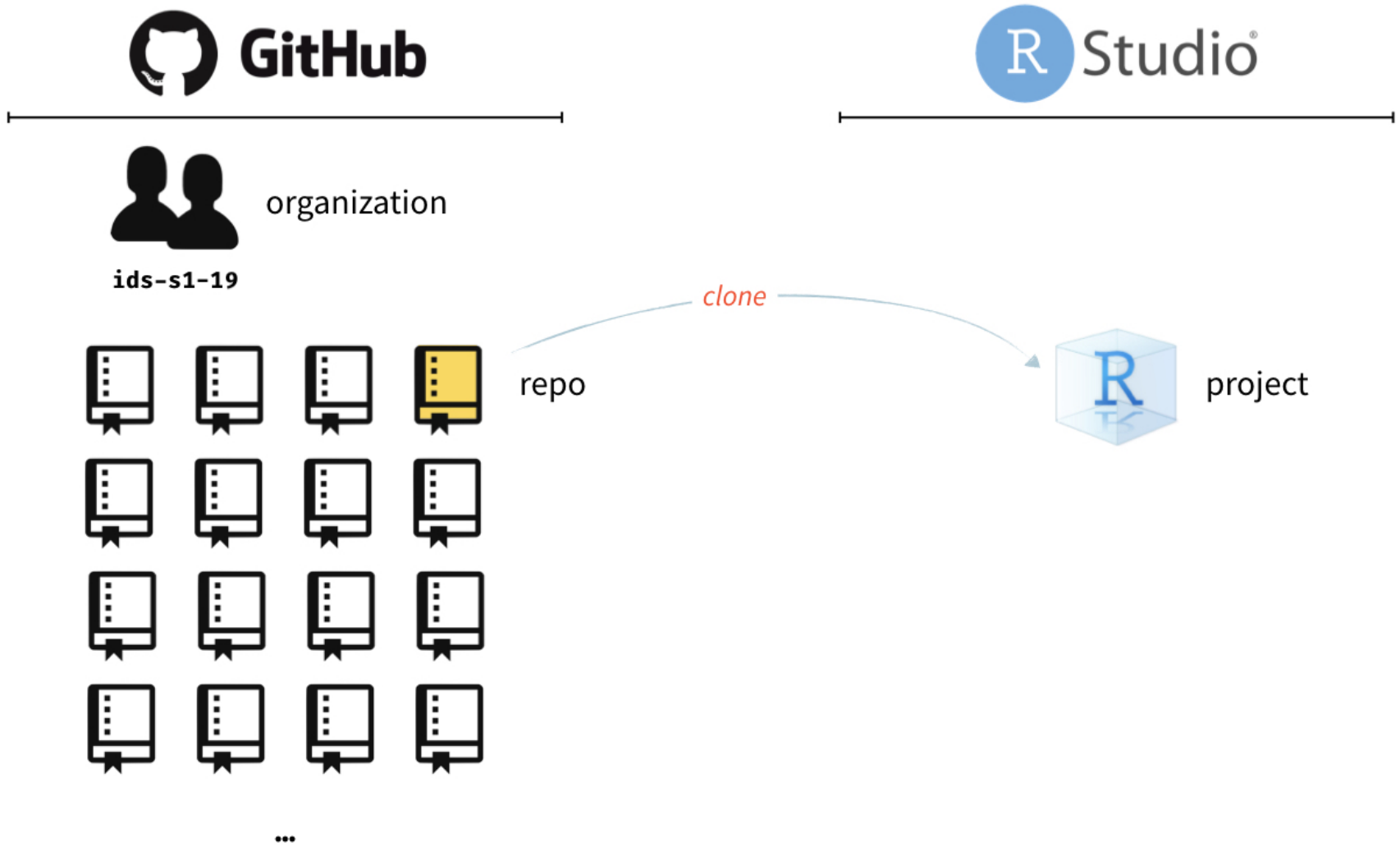
> What is the difference between `.Rmd` and `.md`?

- Usually: `Rmd` -> **knit** -> `.html` or `.pdf`
- But `.html` doesn't show up nice in github
- So we told Rmarkdown to make an `md`, which does show up nice.

🧶 knit
✅ commit
⬆️ push

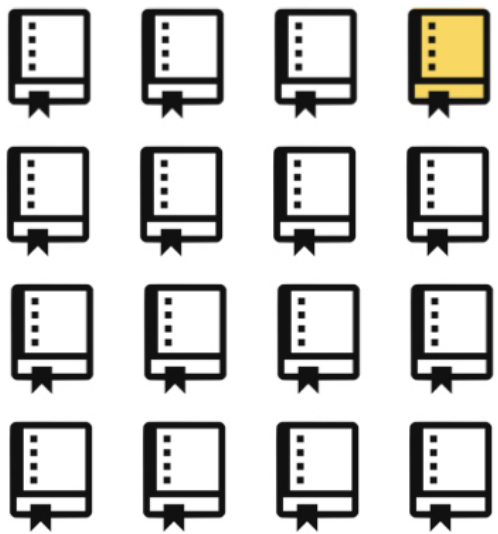

GitHub
organization
ids-s1-19
repo
...


GitHub
organization
ids-s1-19
repo
...

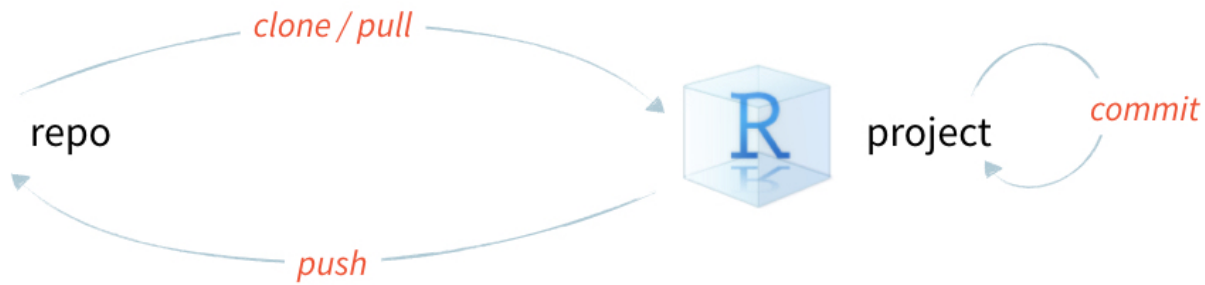

GitHub
RStudio
organization
ids-s1-19
clone
repo
project
...

organization

**ids-s1-19**



...



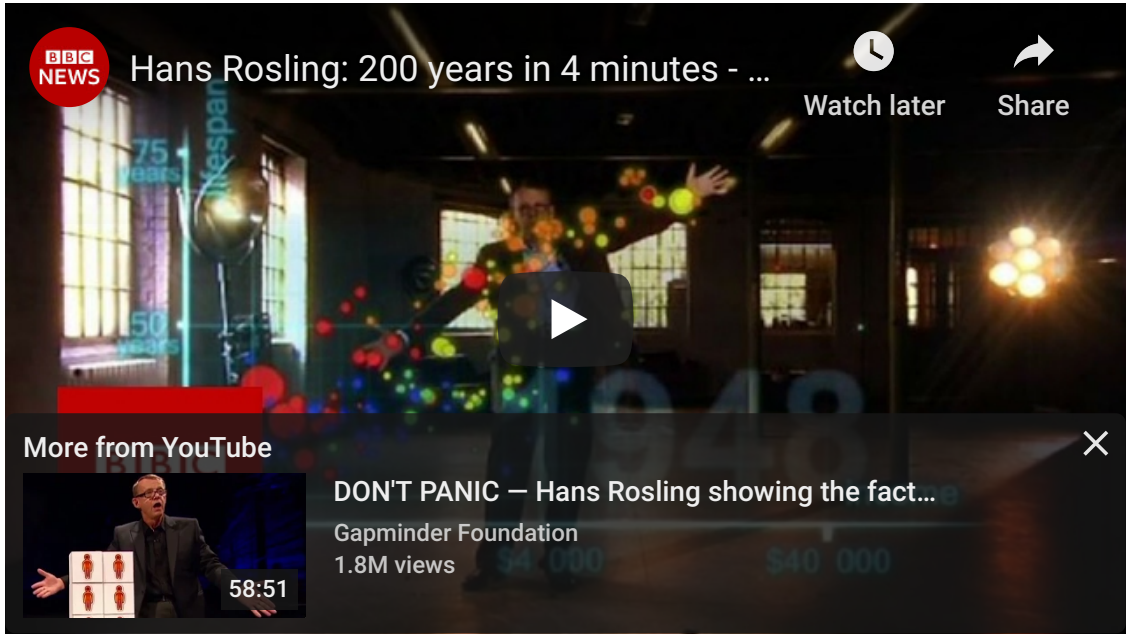
using the R programming language

leveraging functionality provided by R packages

version controlling files with Git

# Asking good questions

- Always include your code and the error
- Create a minimum working example (we'll keep working on this throughout the semester)
- Use code formatting

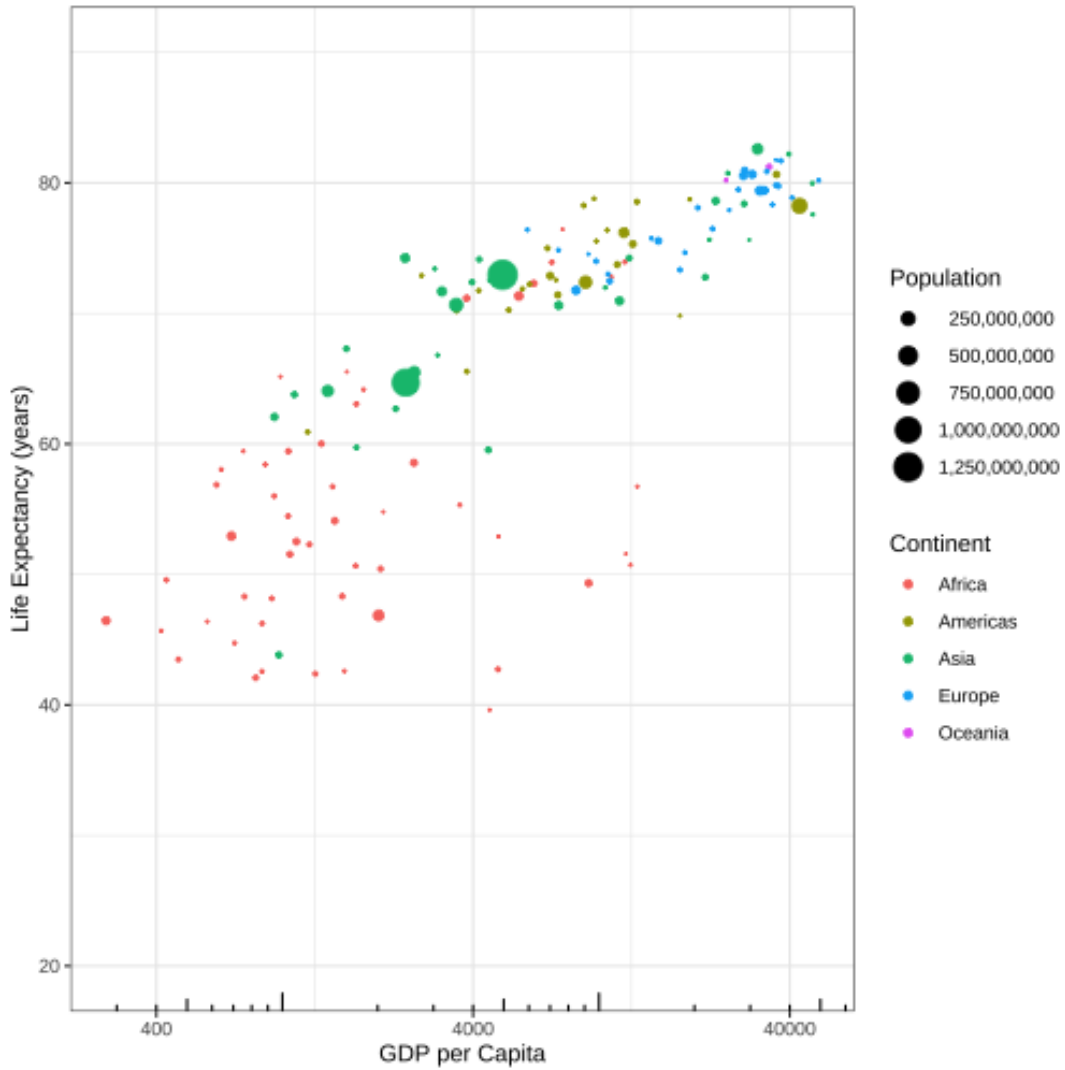

BBC NEWS
Hans Rosling: 200 years in 4 minutes - ...
Watch later
Share
More from YouTube
DON'T PANIC — Hans Rosling showing the fact...
Gapminder Foundation
1.8M views
58:51

```
library(gapminder)
library(ggplot2)
```

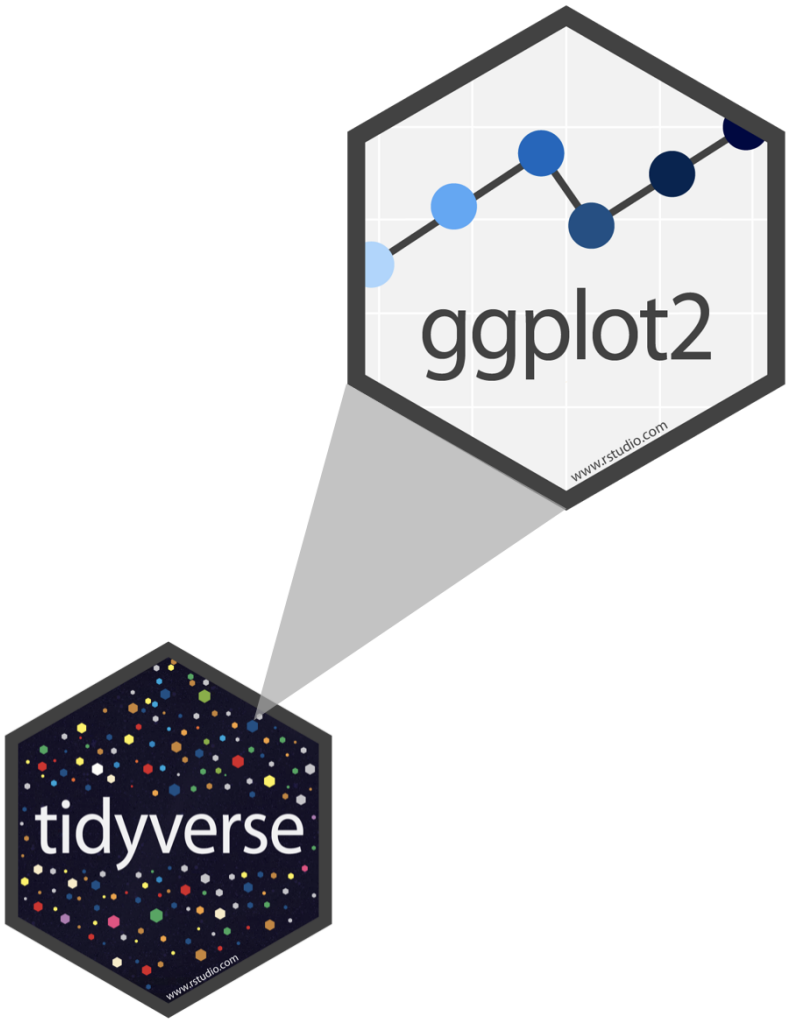
```
gapminder %>% head(5) %>% knitr::kable()
```

| 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 |

# We'll make this chart



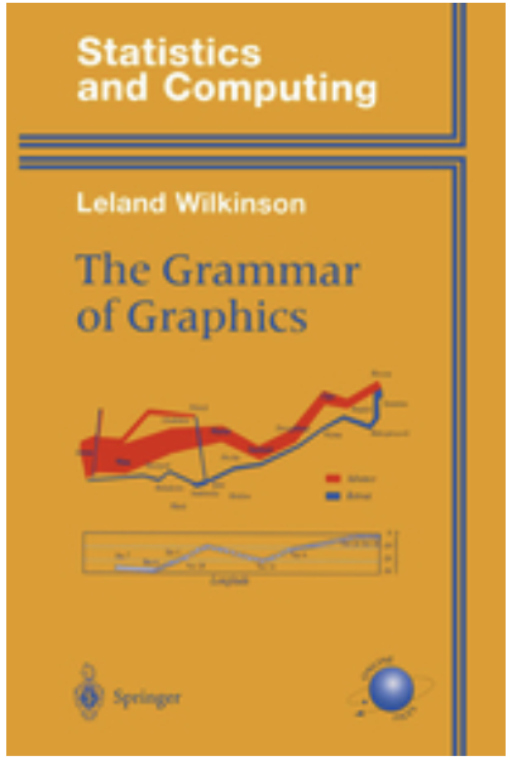
# ggplot2 ∈ tidyverse

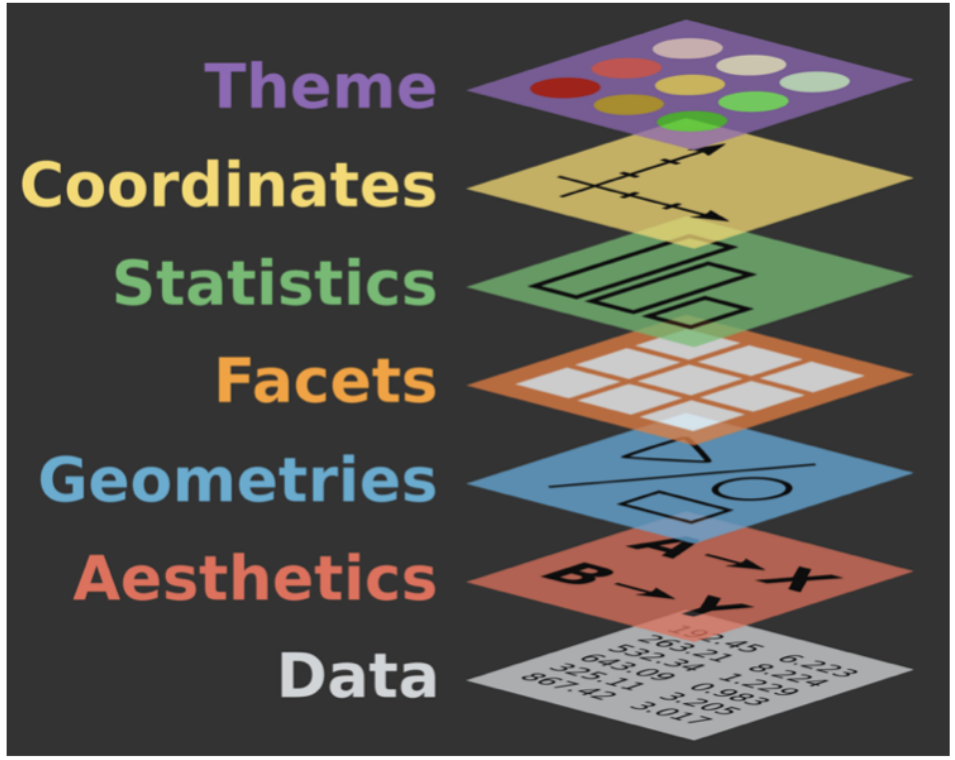



- **ggplot2** is tidyverse's data visualization package
- The gg in "ggplot2" stands for Grammar of Graphics
- It is inspired by the book **Grammar of Graphics** by Leland Wilkinson

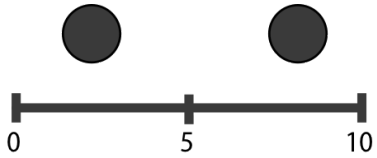
# Grammar of Graphics

Concisely describe the components of a graphic

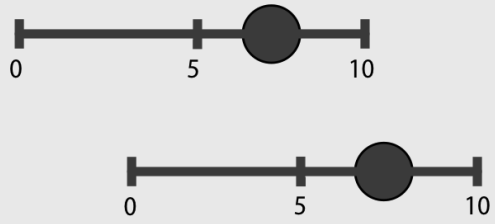



Source: BloggoType



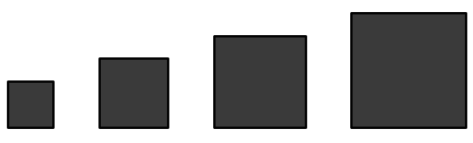

Position on
a common scale



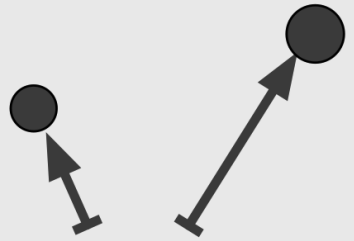

Position on
unaligned
scales

Length

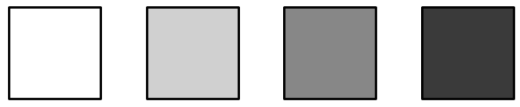
Tilt or Angle



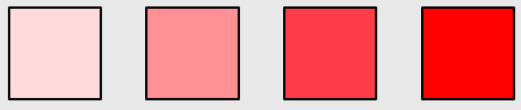
Area (2D as size)



Depth
(3D as position)

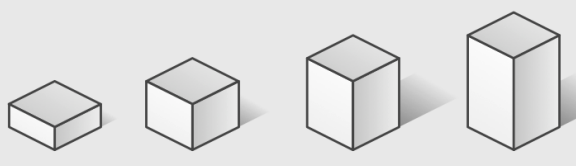


Color luminance
or brightness



Color saturation
or intensity

Curvature



Volume
(3D as size)

Source: https://socviz.co/lookatdata.html#channels-for-representing-data