Week 1
Intro to Tidyverse and ggplot

Data Visualization for Social Scientists

Lecturer: Jeffrey Ziegler, PhD
Teaching Fellow: Shekhar Kedia

ASDS - Trinity College Dublin

Spring 2026

Road map for today

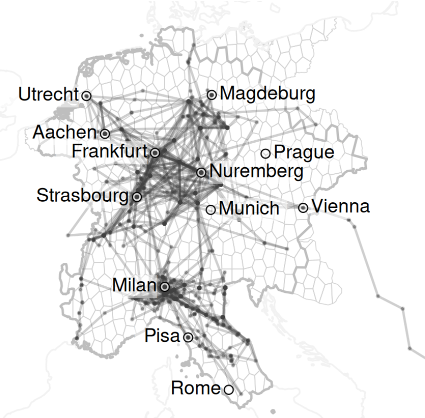
- Today:
  - Welcome
  - Mapping data to aesthetics
  - Intro for tidyverse and ggplot in R
- By next week, please...
  - Fork GitHub repository
  - Problem set #1

Mapping data to aesthetics

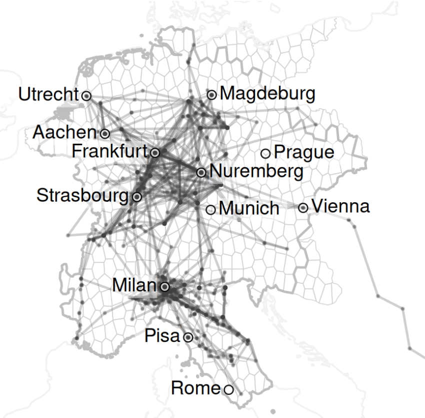
- Aesthetic
  - Visual property of graph
  - Position, shape, color, etc.
- Data
  - Column in dataset

Ex: Rulers' itineraries of Roman Empire

Frederick I Barbarossa (1152-1190)



Louis IV (1314-1347)



Müller-Crepon, C., Neupert-Wentz, C., Kokkonen, A., & Møller, J. (2025). Rulers on the Road: Itinerant Rule in the Holy Roman Empire, AD 919–1519. American Journal of Political Science.

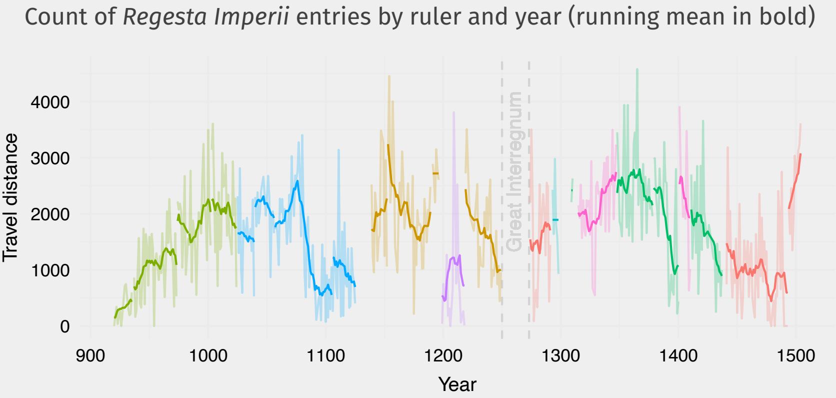
Ex: Mapping data to aesthetics

Data	Aesthetic	Graphic/Geometry
Longitude	Position (x-axis)	Grid
Latitude	Position (y-axis)	Grid
Itinerary	Color	Path
Destination	Color	Point

In ggplot...

Data	aes()	geom
Longitude	x	aes(x=)
Latitude	y	aes(y=)
Itinerary	x, y	geom_line()
Destination	x, y	geom_point()

Ex: Mapping data to aesthetics



Müller-Crepon, C., Neupert-Wentz, C., Kokkonen, A., & Møller, J. (2025). Rulers on the Road: Itinerant Rule in the Holy Roman Empire, AD 919–1519. American Journal of Political Science.

Ex: Mapping data to aesthetics

Data	**aes()**	**geom**
Time	x	aes(x=)
Distance	y	aes(y=)
Leader	x, y	geom_line()
Mean	x, y	geom_line()

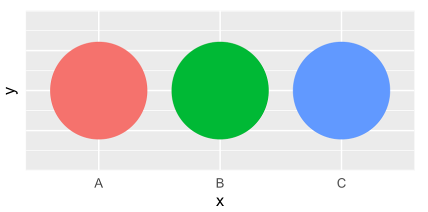
Components as layers

- So far we’ve introduced data, aesthetics, and geometries
- Think of these components as "layers"
- There are many layers we can use
  - We sequentially add layers onto foundational ggplot() plot to create complex figures
  - Add them to initial plot in ggplot() with "+"

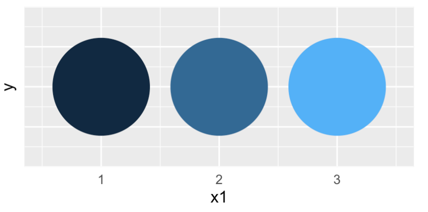
Possible aesthetics

Color - `aes(color=)`

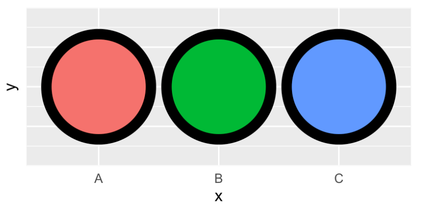
Continuous



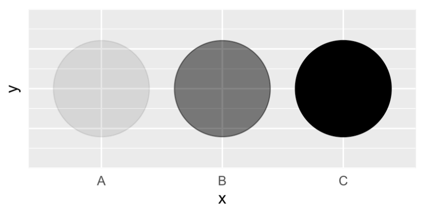
Discrete



Fill - `aes(fill=)`

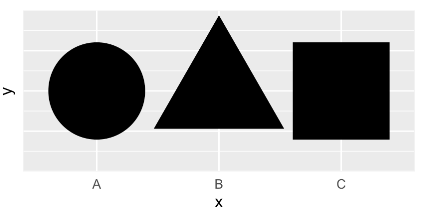


Alpha - `aes(alpha=)`

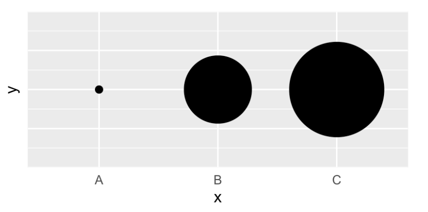


Possible aesthetics

Shape - `geom_point()`



Size - `geom_point(size=)`



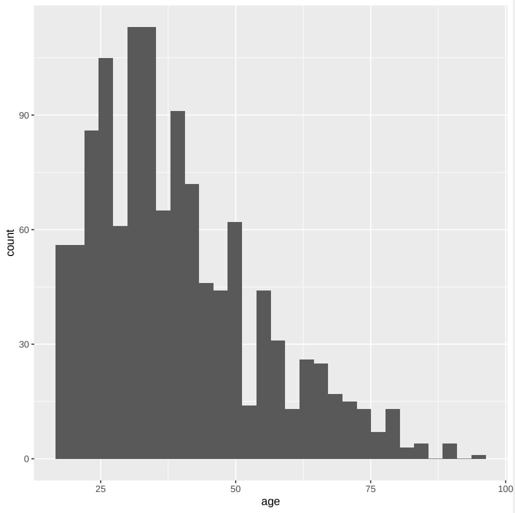
Possible geoms

	Example geom	What it makes
	`geom_col()`	Bar charts
text	`geom_text()`	Text
	`geom_point()`	Points
	`geom_boxplot()`	Boxplots
	`geom_sf()`	Maps

- Dozens of possible geoms, each class session will cover different ones)
- See ggplot2 documentation for complete examples of all geom layers

Ex: Afrobarometer - Zimbabwe (R10, 2024)

Distribution of respondents by age



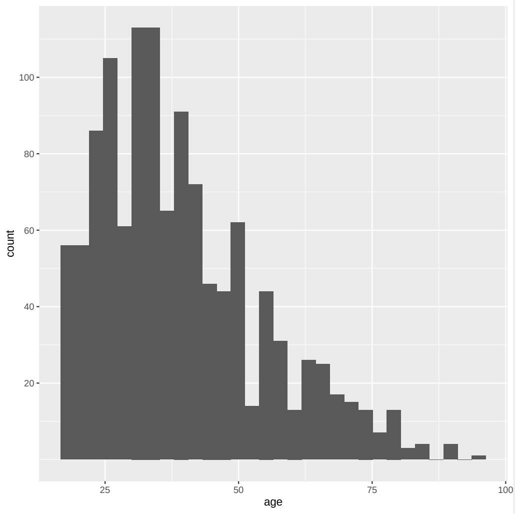
```r
ggplot(data = AB_ZIM,
     aes(x=age)) +
  geom_histogram()
```

Scales

Scales change properties of variable mapping

Example layer	What it does
`scale_x_continuous()`	Make the x-axis continuous
`scale_x_continuous(breaks = 1:5)`	Manually specify axis ticks
`scale_x_log10()`	Log x-axis
`scale_color_gradient()`	Use gradient
`scale_fill_viridis_d()`	Fill with discrete viridis colors

Ex: Adjust y-axis scale



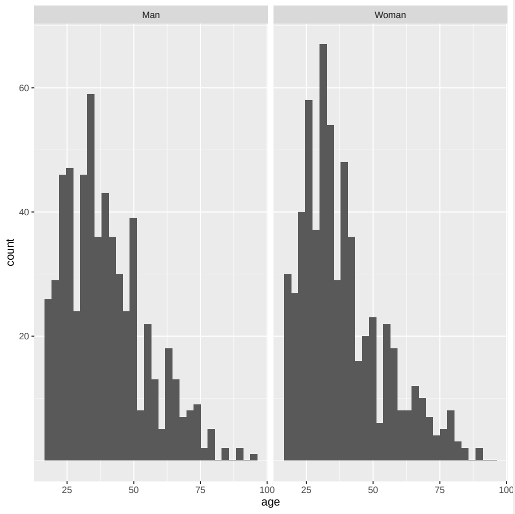
```
ggplot(data = AB_ZIM,
      aes(x=age)) +
  geom_histogram() +
  scale_y_continuous(
     breaks = c(20,
     40, 60, 80, 100))
```

FACETS

Facets show subplots for different subsets of data

Example layer	What it does
`facet_wrap(vars(continent))`	Plot for each continent
`facet_wrap(vars(continent, year))`	Plot for each continent/year
`facet_wrap(..., ncol = 1)`	Put all facets in one column
`facet_wrap(..., nrow = 1)`	Put all facets in one row

Ex: Add facets for gender



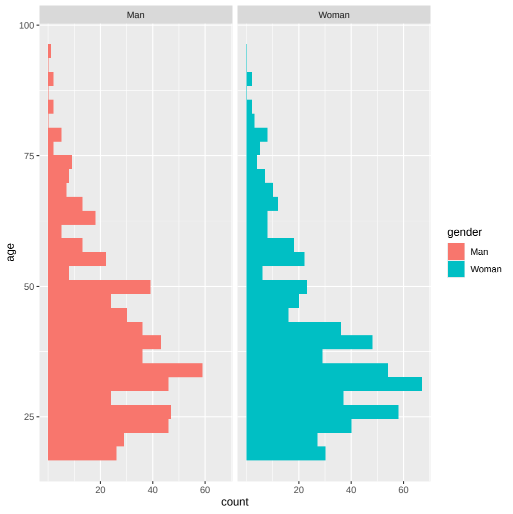
```
ggplot(data = AB_ZIM,
      aes(x=age)) +
  geom_histogram() +
  scale_y_continuous(
    breaks = c(20,
    40, 60, 80, 100))
      +
  facet_wrap(vars(
    gender))
```

Coordinates

Change coordinate system

Example layer	What it does
`coord_cartesian()`	Plot for each continent
`coord_cartesian(ylim = c(1, 10))`	Zoom in where y is 1–10
`coord_flip()`	Switch x and y
`coord_polar()`	Use circular polar system

Ex: Flip coordinates



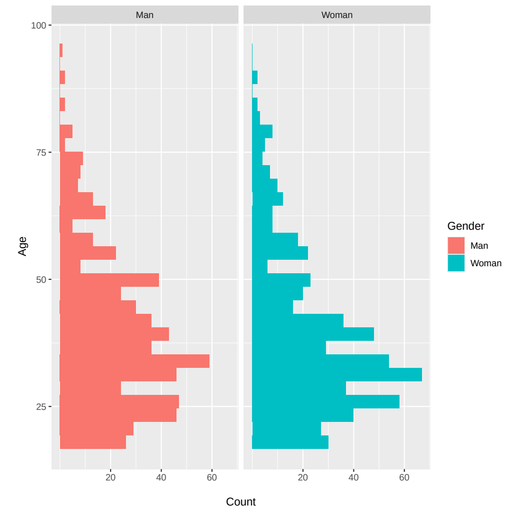
```
ggplot(data = AB_ZIM,
       aes(x=age, fill=
    gender)) +
  geom_histogram() +
   scale_y_continuous(
    breaks = c(20,
    40, 60, 80, 100))
     +
   facet_wrap(vars(
    gender)) +
   coord_flip()
```

Labels

Add labels to plot with a single `labs()` layer

Example layer	What it does
`labs(title = "Neat title")`	Title
`labs(caption = "Something")`	Caption
`labs(y = "Something")`	y-axis
`labs(size = "Population")`	Title of size legend

Ex: Add labels



```
ggplot(data = AB_ZIM, aes(x=age, fill=gender)) +
  geom_histogram() +
  scale_y_continuous(breaks = c(20, 40, 60, 80, 100)) +
  facet_wrap(vars(gender)) +
  coord_flip() +
  labs(x="\nAge", y="\nCount", fill="Gender")
```

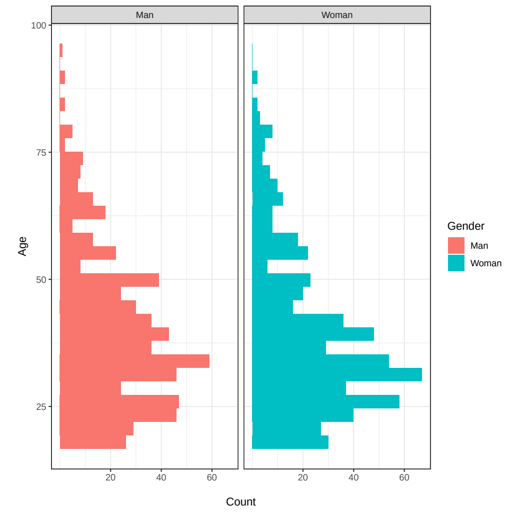
Theme

Change appearance of anything in plot

- Many built-in themes
- Pre-built themes online, like ggthemes package
- Make theme adjustments with `theme()`
- So many options, we have a whole class session dedicated to this

Example layer	**What it does**
`theme_grey()`	Default grey background
`ltheme_bw()`	Black and white
`theme_dark()`	Dark
`theme_minimal()`	Minimal

Ex: Apply theme_bw

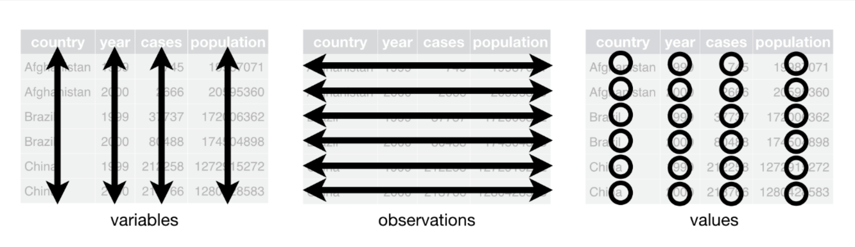


```r
ggplot(data = AB_ZIM,
      aes(x=age, fill=
    gender)) +
  geom_histogram() +
  scale_y_continuous(
    breaks = c(20,
    40, 60, 80, 100))
     +
  facet_wrap(vars(
    gender)) +
  coord_flip() +
  labs(x="\nAge", y="
    \nCount", fill="
    Gender") +
  theme_bw()
```

Cleaning data for plotting - tidyverse

- For `ggplot()` to work, your data needs to be in a tidy format
- This doesn't mean that it's "clean", refers to structure of data
- All the packages in the **tidyverse** work best to tidy data

Visualizing tidy data

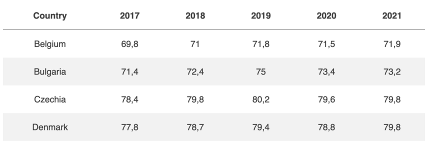


- Each variable has its own column
- Each observation has its own row
- Each value has its own cell

Wide vs. long data

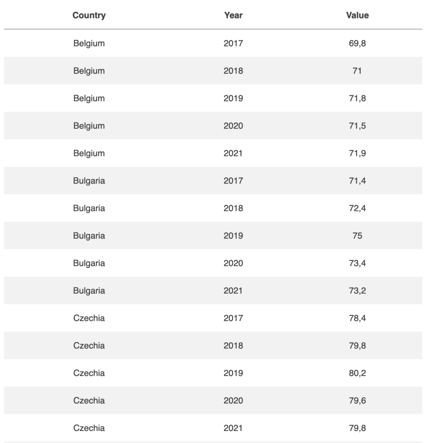
Ex: Employment rate by country

Wide



Country	2017	2018	2019	2020	2021
Belgium	69,8	71	71,8	71,5	71,9
Bulgaria	71,4	72,4	75	73,4	73,2
Czechia	78,4	79,8	80,2	79,6	79,8
Denmark	77,8	78,7	79,4	78,8	79,8

Long



Country	Year	Value
Belgium	2017	69,8
Belgium	2018	71
Belgium	2019	71,8
Belgium	2020	71,5
Belgium	2021	71,9
Bulgaria	2017	71,4
Bulgaria	2018	72,4
Bulgaria	2019	75
Bulgaria	2020	73,4
Bulgaria	2021	73,2
Czechia	2017	78,4
Czechia	2018	79,8
Czechia	2019	80,2
Czechia	2020	79,6
Czechia	2021	79,8

Review

- Mapping data to aesthetics
- ggplot
  - aes()
  - geom
  - Layers
  - Scales
  - Facets
  - Coordinates
  - Labels
  - Themes

Class business

- Read required (and suggested) online materials
- Fork GitHub repository
- Problem set # 1 is up on GitHub