## Introduction to the Grammar of Graphics

Prof. Wells

STA 209, 1/27/23

## Outline

In this lecture, we will...

### Outline

In this lecture, we will...

- Motivate intentional data visualization
- Discuss the Grammar of Graphics
- Decompose particular graphics using the GG paradigm

# Section 1

Data Visualization

• Even small data sets are big. We need to summarize the data.

- Even small data sets are big. We need to summarize the data.
- Humans are best at finding patterns in visual media.

- Even small data sets are big. We need to summarize the data.
- Humans are best at finding patterns in visual media.
- Graphs allow us to compare and explore relationships between variables.

- Even small data sets are big. We need to summarize the data.
- Humans are best at finding patterns in visual media.
- Graphs allow us to compare and explore relationships between variables.
- Most importantly, graphs tell a compelling story.

• In this class, we will encounter a variety of data sets and data contexts

- In this class, we will encounter a variety of data sets and data contexts
  - Some of this data may be light, humorous, or low-stakes.

- In this class, we will encounter a variety of data sets and data contexts
  - Some of this data may be light, humorous, or low-stakes.
- But we encounter data sets and data contexts that are pertinent to our contemporary conversations

- In this class, we will encounter a variety of data sets and data contexts
  - Some of this data may be light, humorous, or low-stakes.
- But we encounter data sets and data contexts that are pertinent to our contemporary conversations
  - Some of this data may be serious, with pertinent and real ramifications

- In this class, we will encounter a variety of data sets and data contexts
  - Some of this data may be light, humorous, or low-stakes.
- But we encounter data sets and data contexts that are pertinent to our contemporary conversations
  - Some of this data may be serious, with pertinent and real ramifications
  - Some data you may have been exposed to before, via news media, social networks, or other discussion

- In this class, we will encounter a variety of data sets and data contexts
  - Some of this data may be light, humorous, or low-stakes.
- But we encounter data sets and data contexts that are pertinent to our contemporary conversations
  - Some of this data may be serious, with pertinent and real ramifications
  - Some data you may have been exposed to before, via news media, social networks, or other discussion
  - Some data may touch on topics are uncomfortable or that may elicit strong emotion.
  - Such data may impact each of us differently, based on our own lived-in experiences.

- In this class, we will encounter a variety of data sets and data contexts
  - Some of this data may be light, humorous, or low-stakes.
- But we encounter data sets and data contexts that are pertinent to our contemporary conversations
  - Some of this data may be serious, with pertinent and real ramifications
  - Some data you may have been exposed to before, via news media, social networks, or other discussion
  - Some data may touch on topics are uncomfortable or that may elicit strong emotion.
  - Such data may impact each of us differently, based on our own lived-in experiences.
- It is important to study data both of serious and light natures, because statistics is an important tool that is used (and misused) for communicating

- In this class, we will encounter a variety of data sets and data contexts
  - Some of this data may be light, humorous, or low-stakes.
- But we encounter data sets and data contexts that are pertinent to our contemporary conversations
  - Some of this data may be serious, with pertinent and real ramifications
  - Some data you may have been exposed to before, via news media, social networks, or other discussion
  - Some data may touch on topics are uncomfortable or that may elicit strong emotion.
  - Such data may impact each of us differently, based on our own lived-in experiences.
- It is important to study data both of serious and light natures, because statistics is an important tool that is used (and misused) for communicating
- My aim is to present data and data contexts with compassionate framing

- In this class, we will encounter a variety of data sets and data contexts
  - Some of this data may be light, humorous, or low-stakes.
- But we encounter data sets and data contexts that are pertinent to our contemporary conversations
  - Some of this data may be serious, with pertinent and real ramifications
  - Some data you may have been exposed to before, via news media, social networks, or other discussion
  - Some data may touch on topics are uncomfortable or that may elicit strong emotion.
  - Such data may impact each of us differently, based on our own lived-in experiences.
- It is important to study data both of serious and light natures, because statistics is an important tool that is used (and misused) for communicating
- My aim is to present data and data contexts with compassionate framing
  - If you have concerns about data and data contexts, please don't hesitate to contact me to share or discuss.

• On January 27th, 1986, engineers from Morton Thiokol recommended NASA delay launch of space shuttle *Challenger* due to cold weather.

- On January 27th, 1986, engineers from Morton Thiokol recommended NASA delay launch of space shuttle *Challenger* due to cold weather.
  - Believed cold weather impacted the o-rings that held the rockets together.
  - Used 13 charts in their argument.

- On January 27th, 1986, engineers from Morton Thiokol recommended NASA delay launch of space shuttle *Challenger* due to cold weather.
  - Believed cold weather impacted the o-rings that held the rockets together.
  - Used 13 charts in their argument.
- After a two hour conference call, the engineer's recommendation was overruled due to lack of persuasive evidence and the launch proceeded.

- On January 27th, 1986, engineers from Morton Thiokol recommended NASA delay launch of space shuttle *Challenger* due to cold weather.
  - Believed cold weather impacted the o-rings that held the rockets together.
  - Used 13 charts in their argument.
- After a two hour conference call, the engineer's recommendation was overruled due to lack of persuasive evidence and the launch proceeded.
- The Challenger exploded 73 seconds into launch.

• Here is one of those charts:

#### • Here is one of those charts:

BLOW BY HISTORY SRM-15 WORST BLOW-R.		HISTORY	OF C (DEGRE	D-RING TO	EMPERATURES
· 2 CASE JOINTS (80°), (110°) APT	MOTOR	MBT	AMB	O-RING	WIND
· MUCH WORSE VISUALLY THAN SRM-22	Dm-+	68	36	47	IO MPH
	Dm-2	76	45	52	10 mpH
SRM 12 BLOW-BY	Qm - 3	72.5	40	48	10 mpH
· 2 CASE JOINTS (30-40°)	Qm - 4	76	48	51	10 MPH
	SRM-15	52	64	53	10 mpH
SRM-13 A, 15, 16A, 18, 23A 24A	5RM-22	77	78	75	10 mpH
O NOZZLE BLOW-BY	5 RM - 25	55	26	29 27	IO MPH 25 MPH

• Here is another of those charts:

Here is another of those charts:



### A Better Graph?

• The following is a graphic created in RStudio from Edward Tufte's data in *Visual Explanations: Images and Quantities, Evidence and Narrative* 

### A Better Graph?

• The following is a graphic created in RStudio from Edward Tufte's data in *Visual Explanations: Images and Quantities, Evidence and Narrative* 



### With Context

• And this graphic further emphasizes the direction of the trend.

### With Context

• And this graphic further emphasizes the direction of the trend.



• The graph of temperature vs o-ring damage tells a clear story: lower launch temperatures are associated with greater o-ring damage.

- The graph of temperature vs o-ring damage tells a clear story: lower launch temperatures are associated with greater o-ring damage.
  - Does this graphic give evidence that low temperatures were the **cause** of o-ring damage, and ultimately, of the challenger disaster?

- The graph of temperature vs o-ring damage tells a clear story: lower launch temperatures are associated with greater o-ring damage.
  - Does this graphic give evidence that low temperatures were the **cause** of o-ring damage, and ultimately, of the challenger disaster?
- Data does not speak for itself. And the data stories we tell depend on which information we include.

- The graph of temperature vs o-ring damage tells a clear story: lower launch temperatures are associated with greater o-ring damage.
  - Does this graphic give evidence that low temperatures were the **cause** of o-ring damage, and ultimately, of the challenger disaster?
- Data does not speak for itself. And the data stories we tell depend on which information we include.
  - What are some other possible reasons why rockets o-ring damage and temperature are associated?

- The graph of temperature vs o-ring damage tells a clear story: lower launch temperatures are associated with greater o-ring damage.
  - Does this graphic give evidence that low temperatures were the **cause** of o-ring damage, and ultimately, of the challenger disaster?
- Data does not speak for itself. And the data stories we tell depend on which information we include.
  - What are some other possible reasons why rockets o-ring damage and temperature are associated?
  - How do we measure the variables included?

- The graph of temperature vs o-ring damage tells a clear story: lower launch temperatures are associated with greater o-ring damage.
  - Does this graphic give evidence that low temperatures were the **cause** of o-ring damage, and ultimately, of the challenger disaster?
- Data does not speak for itself. And the data stories we tell depend on which information we include.
  - What are some other possible reasons why rockets o-ring damage and temperature are associated?
  - How do we measure the variables included?
  - Is there a difference between environmental temperature and o-ring temperature?
  - What type of damage is recorded in the "damage index"?

- The graph of temperature vs o-ring damage tells a clear story: lower launch temperatures are associated with greater o-ring damage.
  - Does this graphic give evidence that low temperatures were the **cause** of o-ring damage, and ultimately, of the challenger disaster?
- Data does not speak for itself. And the data stories we tell depend on which information we include.
  - What are some other possible reasons why rockets o-ring damage and temperature are associated?
  - How do we measure the variables included?
  - Is there a difference between environmental temperature and o-ring temperature?
  - What type of damage is recorded in the "damage index"?
  - Can we quantify how likely it is that this apparent association arose simply due to chance?

# Section 2

# The Grammar of Graphics

### The Guiding Principle

A statistical graphic is a mapping of data variables to aesthetic attributes of geometric objects.

### The Guiding Principle

A statistical graphic is a mapping of data variables to aesthetic attributes of geometric objects.

<u>data</u>	aesthetics	geometric object
Planet Name	x position	bar
Planet Diameter	y height	bar
Planet Name	color	bar

### The Guiding Principle

A statistical graphic is a mapping of data variables to aesthetic attributes of geometric objects.





• data: data frame that contains the raw data and variables of interest

- data: data frame that contains the raw data and variables of interest
- geom: geometric shape that the data are mapped to. Can be points, lines, bars, etc.

- data: data frame that contains the raw data and variables of interest
- geom: geometric shape that the data are mapped to. Can be points, lines, bars, etc.
- **aesthetic**: visual properties of the **geom** object, like *x* position, *y* position, color, fill, shape

- data: data frame that contains the raw data and variables of interest
- geom: geometric shape that the data are mapped to. Can be points, lines, bars, etc.
- **aesthetic**: visual properties of the **geom** object, like *x* position, *y* position, color, fill, shape
- scale: controls how data are mapped to the visual values of the aesthetic. i.e. specifying particular colors or shapes

- data: data frame that contains the raw data and variables of interest
- geom: geometric shape that the data are mapped to. Can be points, lines, bars, etc.
- **aesthetic**: visual properties of the **geom** object, like *x* position, *y* position, color, fill, shape
- scale: controls how data are mapped to the visual values of the aesthetic. i.e. specifying particular colors or shapes
- guide: a legend to help user convert visual display back to the data

### Plotting the Planets

#### Consider the planets data frame, planets\_df:

name	type	diameter	rotation	rings	distance
Mercury	Terrestrial planet	0.382	58.64	FALSE	0.4
Venus	Terrestrial planet	0.949	-243.02	FALSE	0.7
Earth	Terrestrial planet	1.000	1.00	FALSE	1.0
Mars	Terrestrial planet	0.532	1.03	FALSE	1.5
Jupiter	Gas giant	11.209	0.41	TRUE	5.2
Saturn	Gas giant	9.449	0.43	TRUE	9.5
Uranus	Gas giant	4.007	-0.72	TRUE	19.2
Neptune	Gas giant	3.883	0.67	TRUE	30.1

### Plotting the Planets

#### Consider the planets data frame, planets\_df:

name	type	diameter	rotation	rings	distance
Mercury	Terrestrial planet	0.382	58.64	FALSE	0.4
Venus	Terrestrial planet	0.949	-243.02	FALSE	0.7
Earth	Terrestrial planet	1.000	1.00	FALSE	1.0
Mars	Terrestrial planet	0.532	1.03	FALSE	1.5
Jupiter	Gas giant	11.209	0.41	TRUE	5.2
Saturn	Gas giant	9.449	0.43	TRUE	9.5
Uranus	Gas giant	4.007	-0.72	TRUE	19.2
Neptune	Gas giant	3.883	0.67	TRUE	30.1

Describe how to create a plot of distance vs. diameter.

### Plotting the Planets

```
ggplot(data = planets_df, mapping = aes(x = distance, y = diameter)) +
geom_point()
```



### Decomposing Graphics

Let's get some practice decomposing visualizations using the grammar of graphics. For each of images on the next slides, we'll answer the following:

### **Decomposing Graphics**

Let's get some practice decomposing visualizations using the grammar of graphics. For each of images on the next slides, we'll answer the following:

- What is the story this graphic is telling?
- What are the variables here?
- What **geom** are the variables mapped to?
- What are the aesthetics of the geom? Which variable sets the value of that aesthetic?
- What additional context does this graphic provide?

### Example 1 Graphic

#### The Three Types Of Anne Hathaway Movies

Inflation adjusted domestic box office vs. Rotten Tomatoes score



### Example 1

- What is the story this graphic is telling?
- Ø What are the variables here?
- What **geom** are the variables mapped to?
- What are the aesthetics of the geom? Which variable sets the value of that aesthetic?
- 6 What additional context does this graphic provide?



### Example 2 Graphic

#### FiveThirtyEight

¥ f



Polling averages are adjusted based on state and national polls, which means candidates' averages can shift even if no new polls have been added to this page. Read more about the methodology.

### Example 2

- What is the story this graphic is telling?
- Ø What are the variables here?
- What **geom** are the variables mapped to?
- What are the aesthetics of the geom? Which variable sets the value of that aesthetic?
- 6 What additional context does this graphic provide?



oiling averages are adjusted based on state and national polis, which means candidates' averages can shift even if no new polis have been added to this page. Readnore about the methodology.

• Determine what story your graphic should tell. And ensure that all aspects of your graphic serve this purpose.

- Determine what story your graphic should tell. And ensure that all aspects of your graphic serve this purpose.
- Label axes and units for improved clarity.

- Determine what story your graphic should tell. And ensure that all aspects of your graphic serve this purpose.
- Label axes and units for improved clarity.
- Include a legend to translate from aesthetics to variables.

- Determine what story your graphic should tell. And ensure that all aspects of your graphic serve this purpose.
- Label axes and units for improved clarity.
- Include a legend to translate from aesthetics to variables.
- Specify your data source for reproducibility/verification.

- Determine what story your graphic should tell. And ensure that all aspects of your graphic serve this purpose.
- Label axes and units for improved clarity.
- Include a legend to translate from aesthetics to variables.
- Specify your data source for reproducibility/verification.
- Minimize/eliminate extraneous elements that do not serve main purpose.