Lecture 10:
Scatter Plots and Color

April 10, 2023

Course Administration

1. Looking forward
   - Lecture 11: consultations tomorrow, no live lecture
   - Lecture 12: storytelling, and how to make a quarto webpage
   - Lecture 13, Monday May 1: presentations
   - Lecture 14, **Wednesday** May 3: presentations

Course Administration

1. Looking forward
   - Lecture 11: consultations tomorrow, no live lecture
   - Lecture 12: storytelling, and how to make a quarto webpage
   - Lecture 13, Monday May 1: presentations
   - Lecture 14, **Wednesday** May 3: presentations
2. Edited presentation instructions: **presentations due two hours in advance of class**

Course Administration

1. Looking forward
   - Lecture 11: consultations tomorrow, no live lecture
   - Lecture 12: storytelling, and how to make a quarto webpage
   - Lecture 13, Monday May 1: presentations
   - Lecture 14, **Wednesday** May 3: presentations
2. Edited presentation instructions: **presentations due two hours in advance of class**
3. Paper due Monday May 8 by midnight to google drive. **Do not be late.**
4. Anything else?

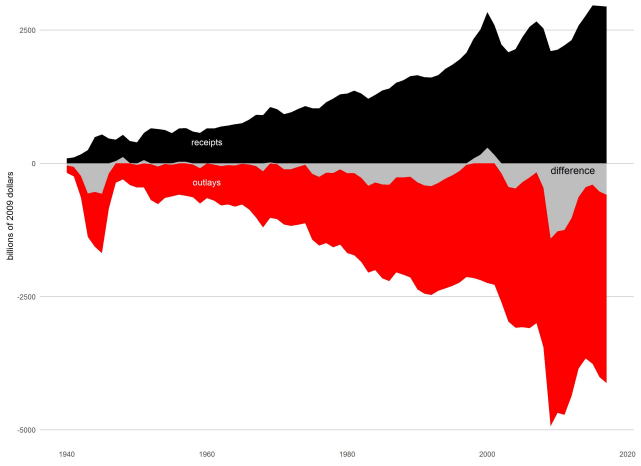
Next Week's Good Bad on Scatters, April 24

Finder	Commenter
Megan	Claire
Hannah	Morgan

This is the last one.

My Surplus Chart

My Surplus Chart



This Lecture

1. Scatter plot definition and origins
2. How and when to use scatters
3. Small multiples
4. Color
5. R stuff

Scatter Plot: Definition and Origins

What is a Scatter Plot?

What is a Scatter Plot?

- Plots values of two different variables on the same chart

What is a Scatter Plot?

- Plots values of two different variables on the same chart
- Shows correlation between two variables

What is a Scatter Plot?

- Plots values of two different variables on the same chart
- Shows correlation between two variables
- Can also show distribution of each variable

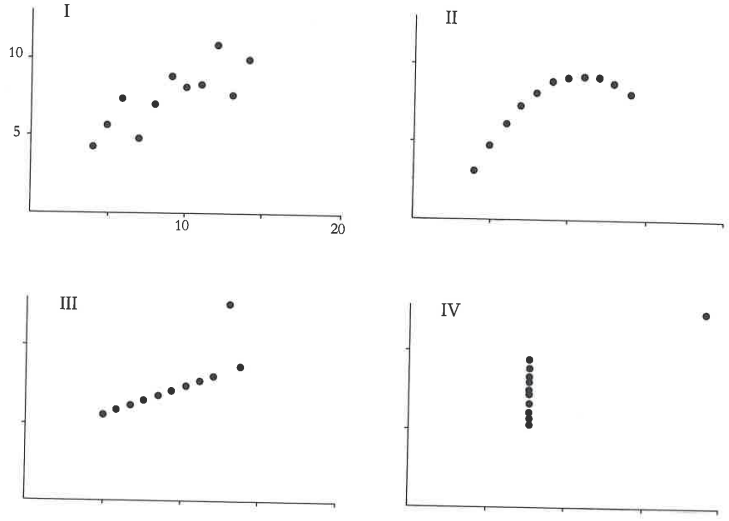
A Reminder and Example: Anscombe's Quartet

Same mean, same variance

I		II		III		IV	
X	Y	X	Y	X	Y	X	Y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

A Reminder and Example: Anscombe's Quartet

Same mean, same variance



What Makes a Scatter Plot Different From All Other Plots?

(That We have Studied) – from Friendly and Denis, 2005

What Makes a Scatter Plot Different From All Other Plots?

(That We have Studied) – from Friendly and Denis, 2005

- It is fundamentally 2-D

What Makes a Scatter Plot Different From All Other Plots?

(That We have Studied) – from Friendly and Denis, 2005

- It is fundamentally 2-D
- A line graph is sort of 2-D, but only really for time

What Makes a Scatter Plot Different From All Other Plots?

(That We have Studied) – from Friendly and Denis, 2005

- It is fundamentally 2-D
- A line graph is sort of 2-D, but only really for time
- Everything else we've studied is either a categorical relationship
  - bar chart

What Makes a Scatter Plot Different From All Other Plots?

(That We have Studied) – from Friendly and Denis, 2005

- It is fundamentally 2-D
- A line graph is sort of 2-D, but only really for time
- Everything else we've studied is either a categorical relationship
  - bar chart
- or 1-D
  - histogram

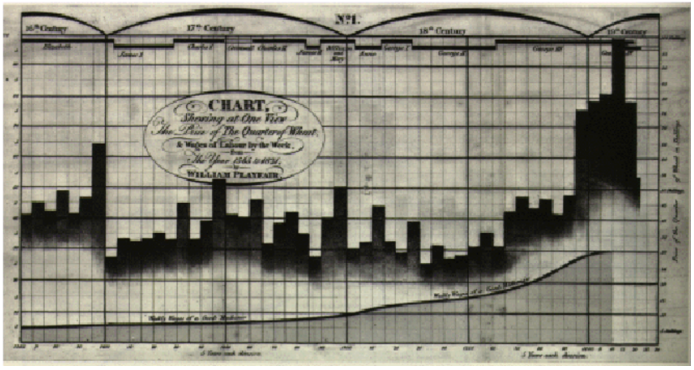
What Makes a Scatter Plot Different From All Other Plots?

(That We have Studied) – from Friendly and Denis, 2005

- It is fundamentally 2-D
- A line graph is sort of 2-D, but only really for time
- Everything else we've studied is either a categorical relationship
  - bar chart
- or 1-D
  - histogram

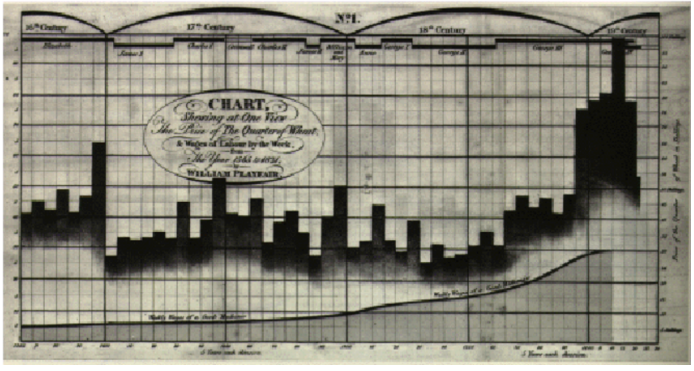
Map is the closest analogue to a scatter: points in (x, y) space

Scatters Are the Most Modern of Graphs We Study



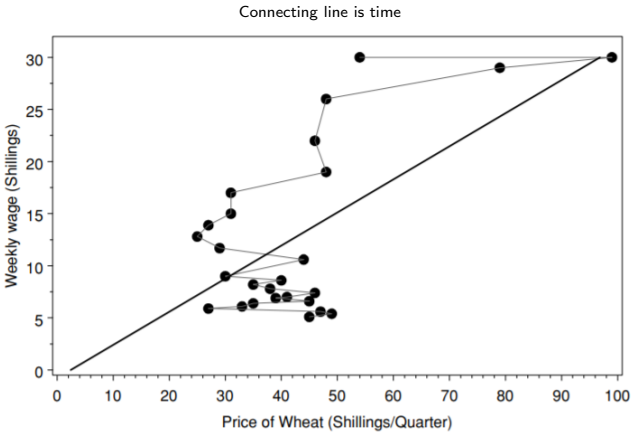
- What is this graph?
  - two y axes
  - wages in line
  - price of wheat in bars
  - horizontal axis is time
- What is the goal of this graph?

Scatters Are the Most Modern of Graphs We Study



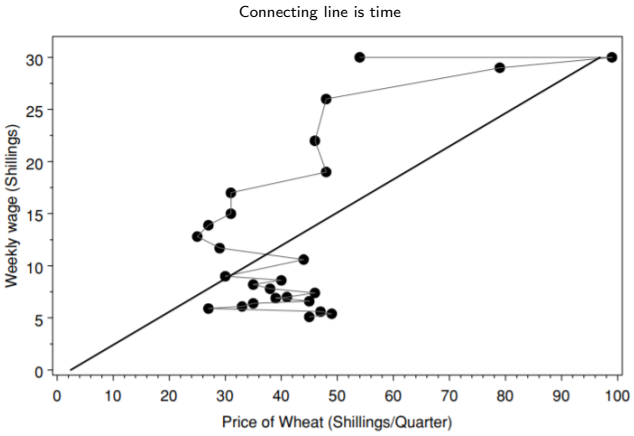
- What is this graph?
  - two y axes
  - wages in line
  - price of wheat in bars
  - horizontal axis is time
- What is the goal of this graph?
  - show that purchasing power increases over time
  - is it clear?

Playfair's Graph as a Proper Scatter



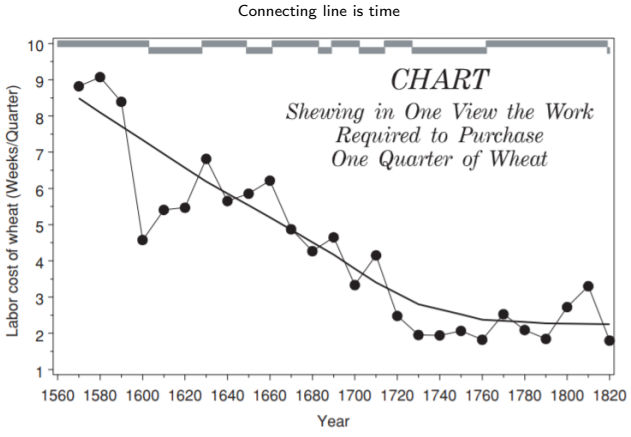
- What is this graph?
  - price of wheat on x
  - wage on y
  - line connects by time

Playfair's Graph as a Proper Scatter



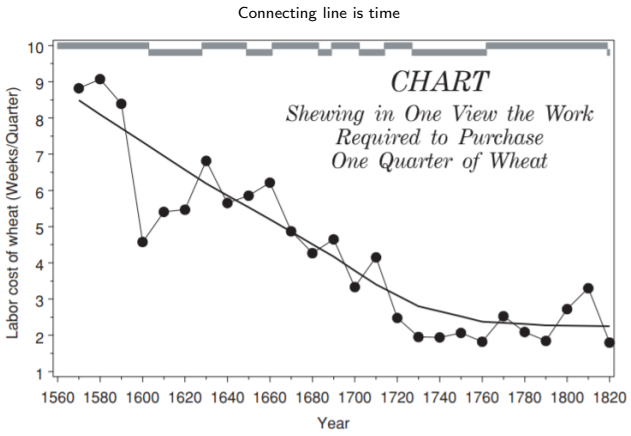
- What is this graph?
  - price of wheat on x
  - wage on y
  - line connects by time
- Why is this graph not too helpful?
  - you don't know when is when
  - no temporal point

Revision of Playfair Makes the Key Point – But is Not a Scatter



- What is this graph?
  - time on x
  - on y, number of weeks required to purchase one quarter of wheat
  - line connects by time

Revision of Playfair Makes the Key Point – But is Not a Scatter



- What is this graph?
  - time on x
  - on y, number of weeks required to purchase one quarter of wheat
  - line connects by time
- Why is this better?
  - line connects time and you can see it
  - makes the ratio for you
  - the ratio is the point!

One of the First Scatterplots: 1886

The Graph

- aims to predict one variable from the other
- has no time dimension
- notes density of observations

One of the First Scatterplots: 1886

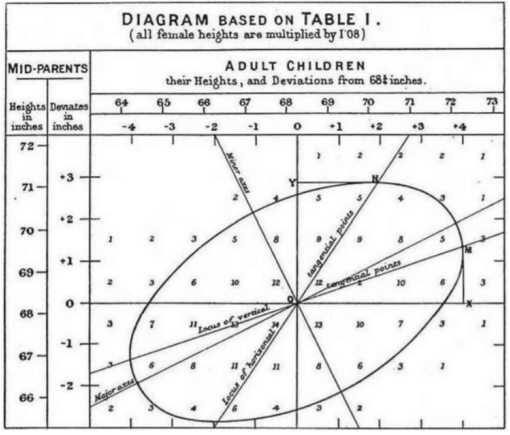
The Graph

- aims to predict one variable from the other
- has no time dimension
- notes density of observations

The Author: Francis Galton

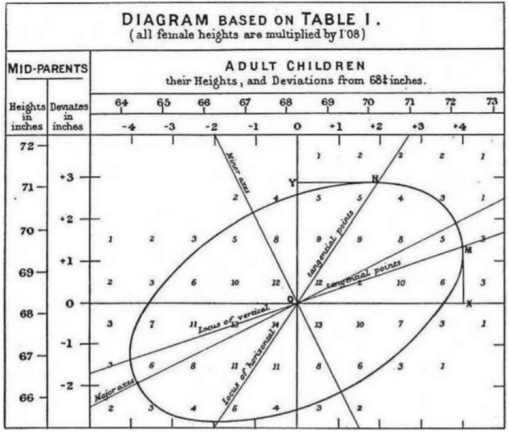
- a measurer of all things: weather, height, etc
- invented or first described
  - the questionnaire
  - standard deviation
  - regression to the mean
- and the developer of eugenics

Galton's Scatter



- What is this graph?
  - height of adult children on x
  - height of parents on y
  - numbers are the number of observations at each point

Galton's Scatter



- What is this graph?
  - height of adult children on x
  - height of parents on y
  - numbers are the number of observations at each point
- This is an early scatter
- Scatters are nor prevalent until the 1920s
- Still usually too complicated for most layperson communications

Galton, 1886.

How and When to Use Scatters

Pros and Cons of Scatters

Most common type of graph for academic presentation

Pros and Cons of Scatters

Most common type of graph for academic presentation

Pros

- Can clearly and compellingly show a bivariate relationship
- Shows relationship throughout the distribution

Pros and Cons of Scatters

Most common type of graph for academic presentation

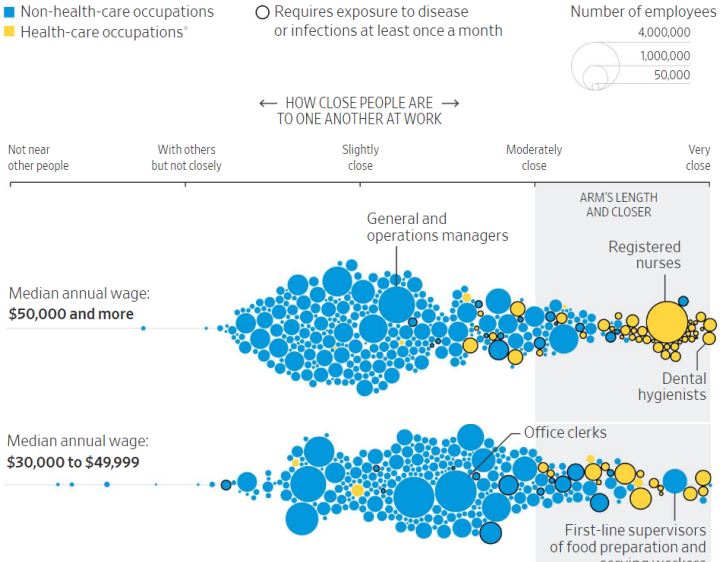
Pros

- Can clearly and compellingly show a bivariate relationship
- Shows relationship throughout the distribution

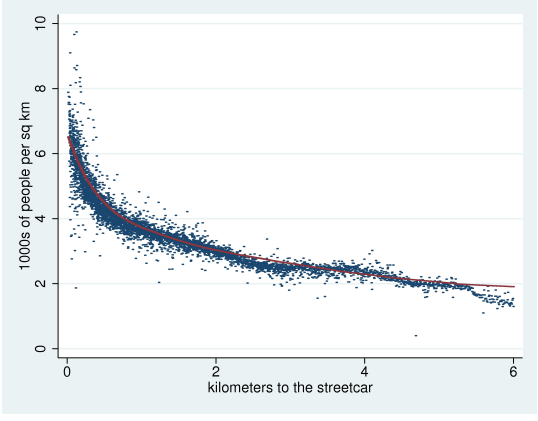
Cons

- Requires the audience to think about the relationship
- Sometimes too complicated for policy communication
- Can obscure relationships that do exist

This Should be a Scatter But Was Not



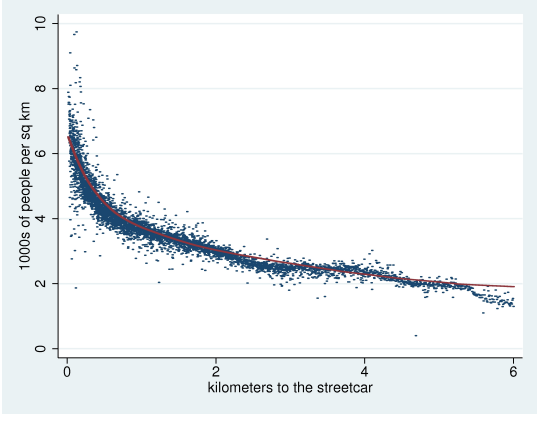
My Best Ever Scatter



What is it?

- Each point is
- average population density near about 400 land plots
- at a given distance from an old streetcar
- red line is a flexible regression line

My Best Ever Scatter

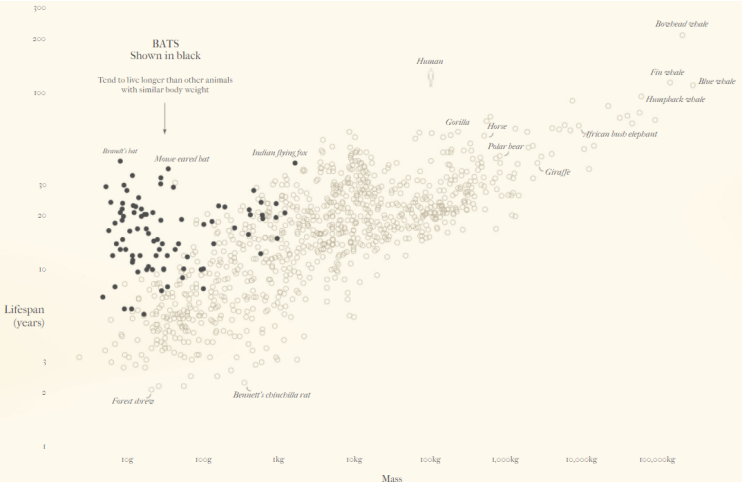


What is it?

- Each point is
- average population density near about 400 land plots
- at a given distance from an old streetcar
- red line is a flexible regression line

Data show the point

How Can You Annotate a Scatter?



- best fit lines
- ovals
- colors
- call out individual items

Showing Multiple Variables or Variations

How to Deal with Issues of Multiple Variables

1. If they are in the same units?

How to Deal with Issues of Multiple Variables

1. If they are in the same units? graph on the same scale

How to Deal with Issues of Multiple Variables

1. If they are in the same units? graph on the same scale
2. If they are in different units?

How to Deal with Issues of Multiple Variables

1. If they are in the same units? graph on the same scale
2. If they are in different units?
   - can use two axes, but rarely a good idea – why?

How to Deal with Issues of Multiple Variables

1. If they are in the same units? graph on the same scale
2. If they are in different units?
   - can use two axes, but rarely a good idea – why?
   - plot on two charts side-by-side
   - do you want side-by-side vertical or horizontal?

How to Deal with Issues of Multiple Variables

1. If they are in the same units? graph on the same scale
2. If they are in different units?
   - can use two axes, but rarely a good idea – why?
   - plot on two charts side-by-side
   - do you want side-by-side vertical or horizontal?
3. If you have many different variables to show?
   - see the next slide..

Small Multiples

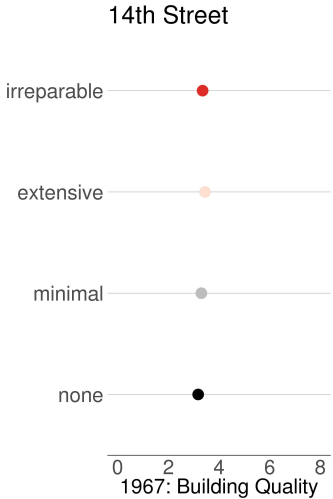
When do you use them?

- Multiple variables to show
- Too much for one graph
- In presentations, usually helpful to explain one part first

There is an implicit assumption that all graphs use the same scale.

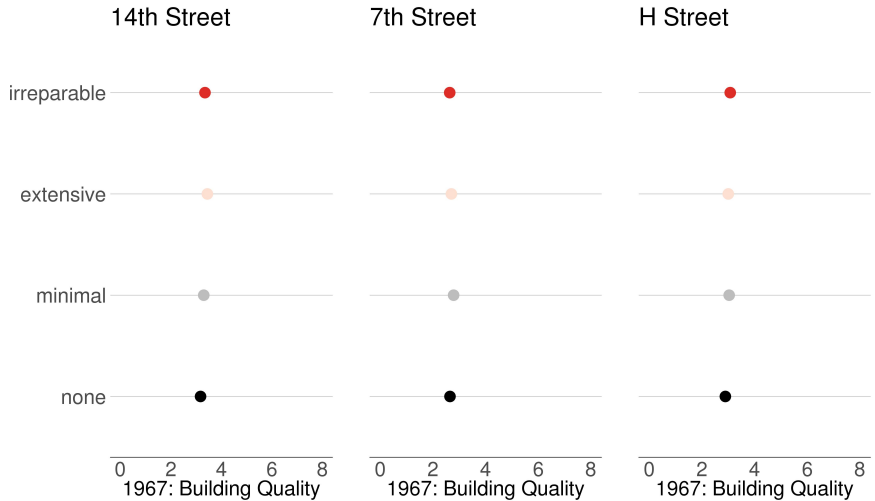
My Small Multiples

Destruction Roughly Even by 1967 Quality



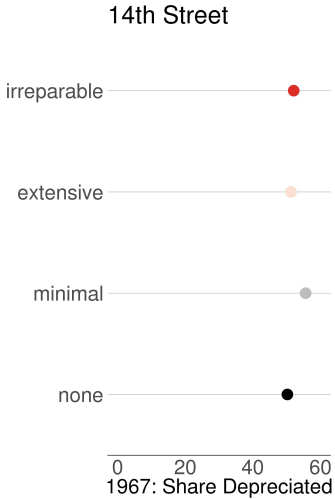
My Small Multiples

Destruction Roughly Even by 1967 Quality



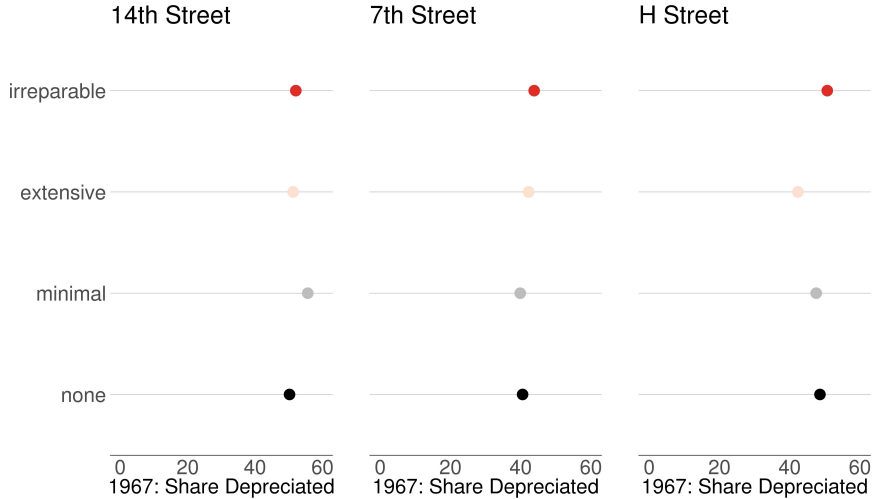
My Small Multiples

Destruction Roughly Even by 1967 Depreciation



My Small Multiples

Destruction Roughly Even by 1967 Depreciation



How Beyonce Exploits the Power of Small Multiples

With thanks to Vibe.

Using Color Well

Color Rules, 1 of 2

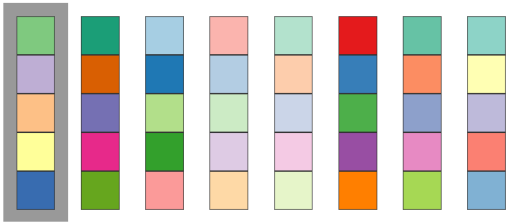
1. Use color because it may be the fastest discriminator
2. Use color because color builds in size and emotion
3. Color should have a function, not be a decoration
4. (repeat) We can't remember too many categories $\rightarrow$ too many colors
5. Things that are the same color are linked, whether you intend to or not
6. Be consistent with color across graphics

Color Rules, 2 of 2

7. Categorical things must get qualitative scales
8. Consecutive continuous things get sequential color scales
9. We think darker = denser → darker = larger → make bigger values darker colors
10. Consecutive continuous things with two binary options can get diverging sequential color scales
11. Use a tool to choose color-blind accessible options
12. All kinds of ways to choose: colorbrewer2.org, metbrewer

With thanks to Cynthia Brewer, Towards Data Science, datawrapper.de, and this Adobe blog.

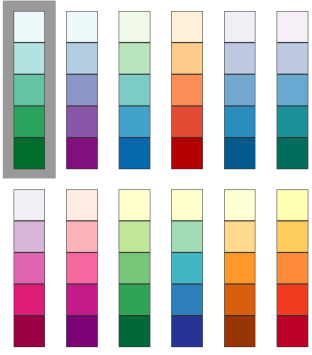
7. Categorial Things Get Qualitative Scales



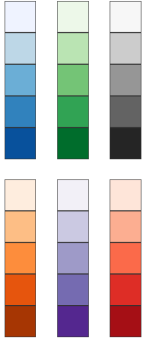
What kind of categorical things would work well here?

8. Sequential Color for Consecutive Continuous Things

Multi-hue:

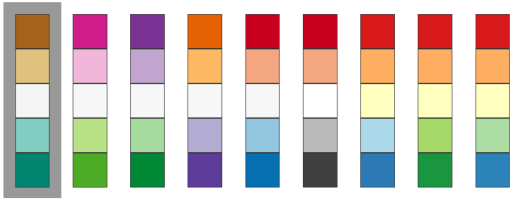


Single hue:



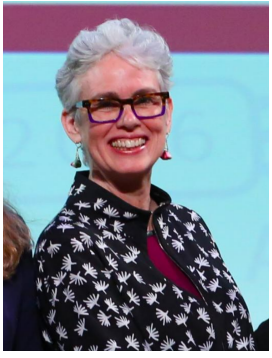
What kind of sequential things would work well here?

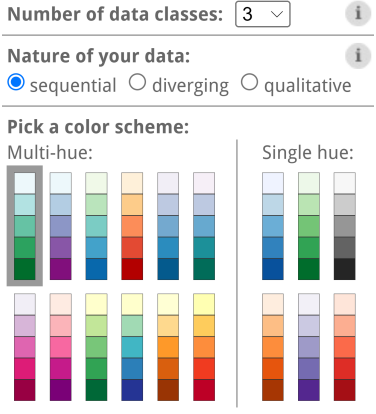
10. Diverging Sequential Color



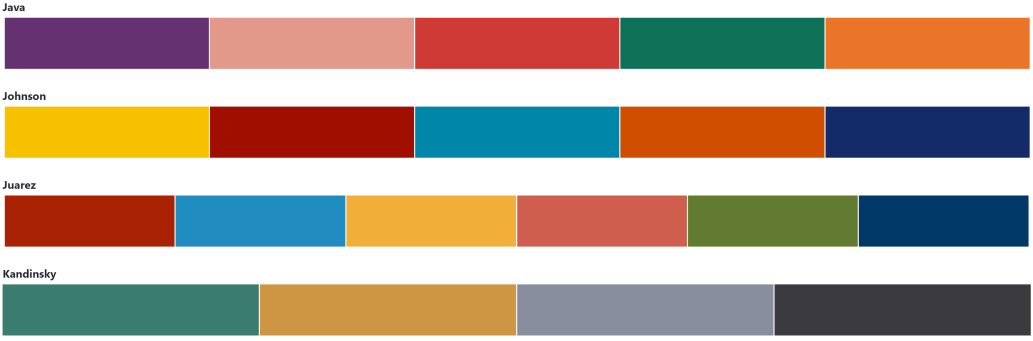
What kind of type of series would work well here?

12. Use Colorbrewer2.org





12, cont'd. MetBrewer Uses This Framework



R Notes on Scatters

Next Lectures

- Consultations this week
- No lecture next week
- Presentations due online 48 hours before you present in class – May 1 or 3
- Final paper due May 8 by midnight