

Lecture 9: Scatter Plots and Color

March 30, 2026

Course Administration

① Looking forward

- Lecture 10: guest speaker, in-class workshop
- Lecture 11: no class, consultations
- Lecture 12: Storytelling & Interactivity
- Lecture 13, 14: Monday and **Wednesday**: presentations

Course Administration

① Looking forward

- Lecture 10: guest speaker, in-class workshop
- Lecture 11: no class, consultations
- Lecture 12: Storytelling & Interactivity
- Lecture 13, 14: Monday and **Wednesday**: presentations

② Presentations due two hours in advance of class

③ Final policy brief due Monday May 4 by midnight. **Do not be late.**

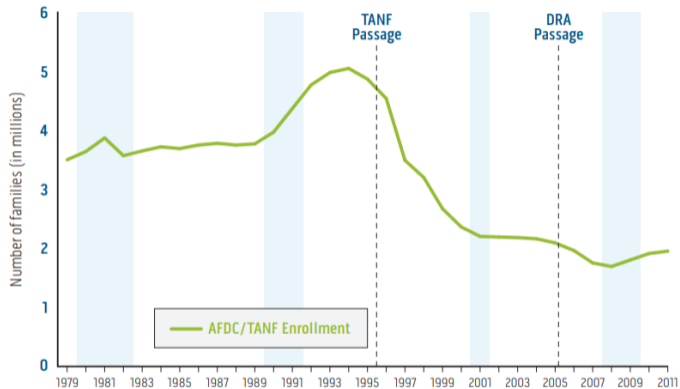
④ Anything else?

Next Week's Final Good Bad Ugly on Scatters

Finder	Commenter
DelRoy	Manavi
Manavi	Thomas

This is the last one.

DelRoy on Eva's Graph

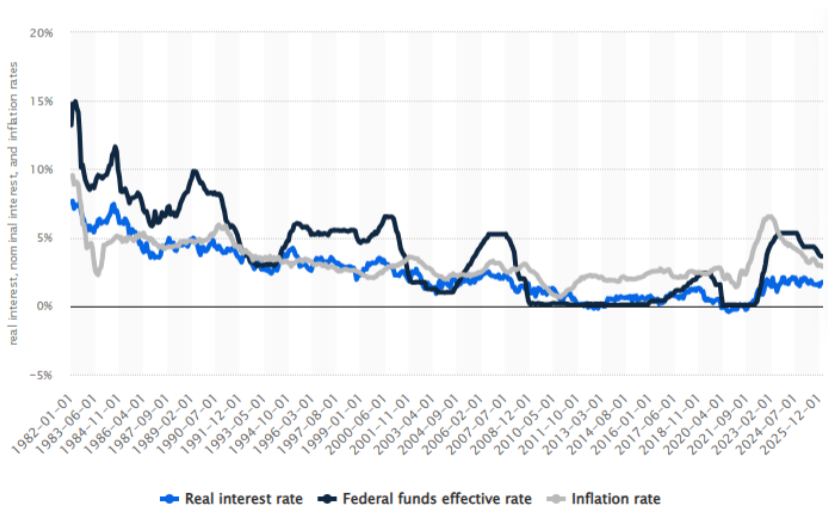


Sources: Administration for Children and Families (TANF/SSP/AFDC enrollment),
National Bureau of Economic Research (recessions).

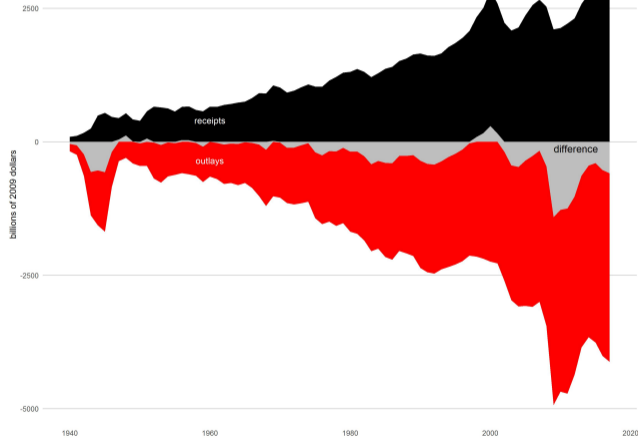
Loprest, Pamela. "How Has the TANF Caseload Changed Over Time?" Urban Institute Brief No. 08, March 2012. [\[link\]](#)

Tosha on Otsile's Graphs

Monthly 10-year real interest rate, Federal Funds effective rate and inflation rate in the United States from 1982 to 2026



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

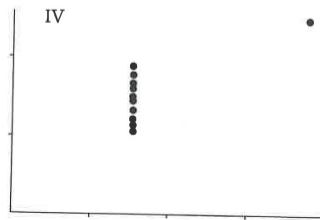
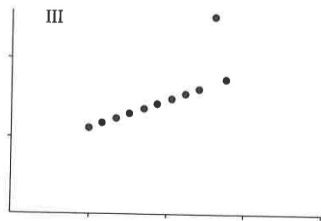
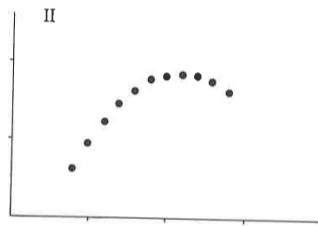
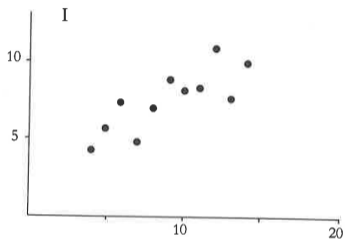
A Reminder and Example: Anscombe's Quartet

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

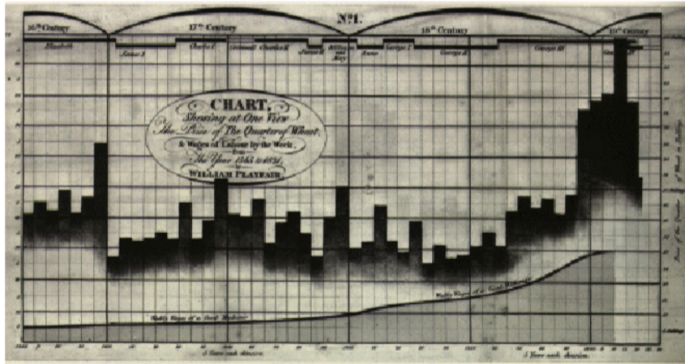
What Makes a Scatter Plot Different From All Other Plots?

(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

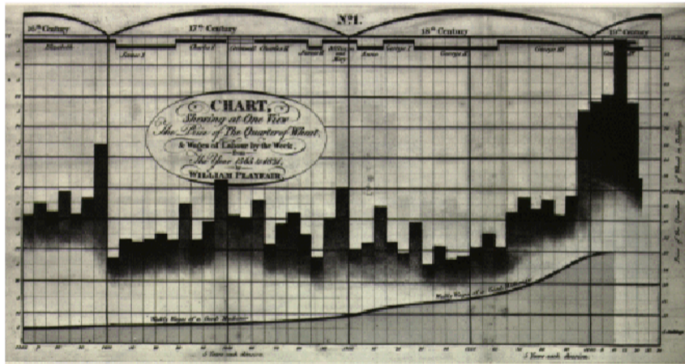
Map is the closest analogue to a scatter: points in (x, y) space

Scatters Are the Most Modern of Graphs We Study



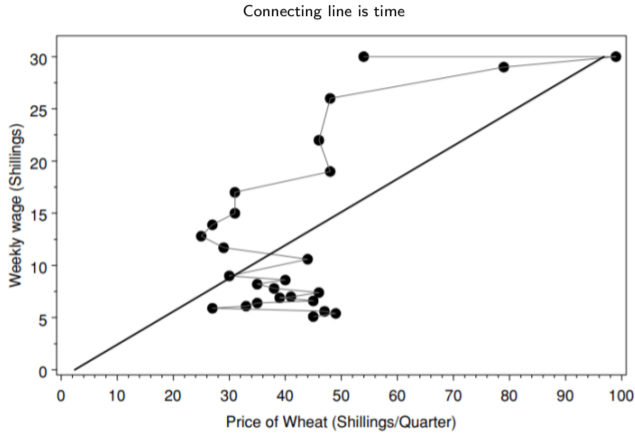
- What is this graph?
 - two y axes
 - wages in line
 - price of wheat in bars
 - horizontal axis is time
- What is the goal of this graph?

Scatters Are the Most Modern of Graphs We Study



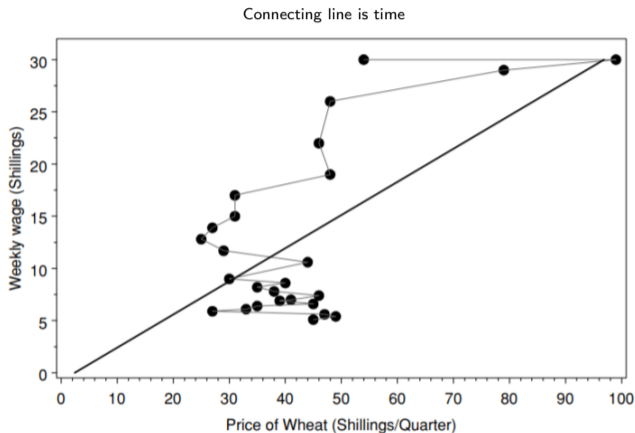
- What is this graph?
 - two y axes
 - wages in line
 - price of wheat in bars
 - horizontal axis is time
- What is the goal of this graph?
 - show that purchasing power increases over time
 - is it clear?

Playfair's Graph as a Proper Scatter



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

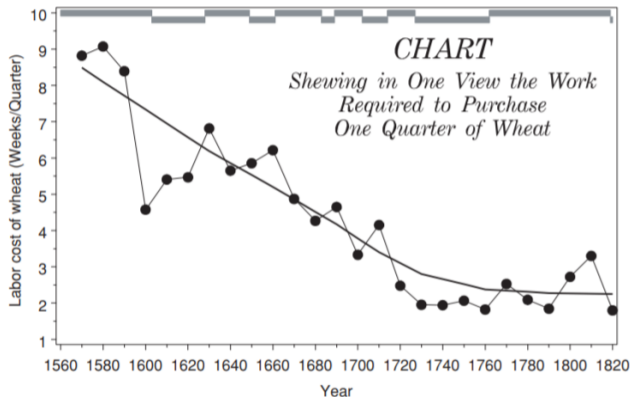
Playfair's Graph as a Proper Scatter



- 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

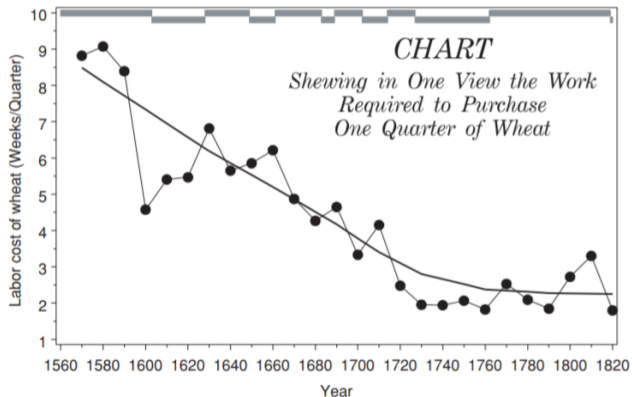
Connecting line is time



- 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

Connecting line is time



- 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

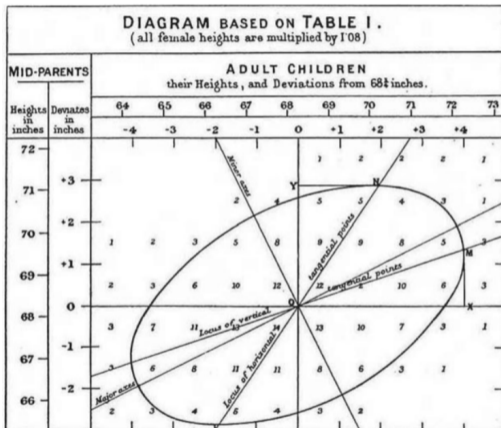
The Graph

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

The Author: Francis Galton

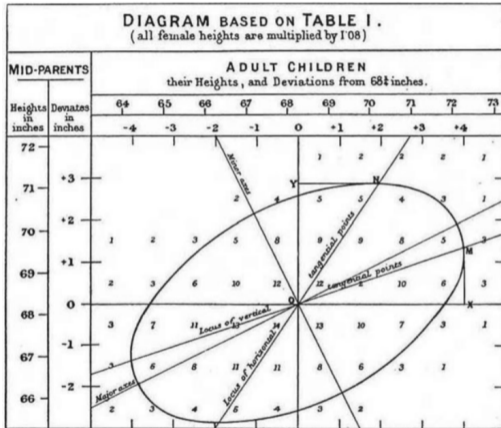
- a measurer of all things: weather, height, etc
- invented or first described
 - the questionnaire
 - standard deviation
 - regression to the mean
- and the developer of eugenics

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 not 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

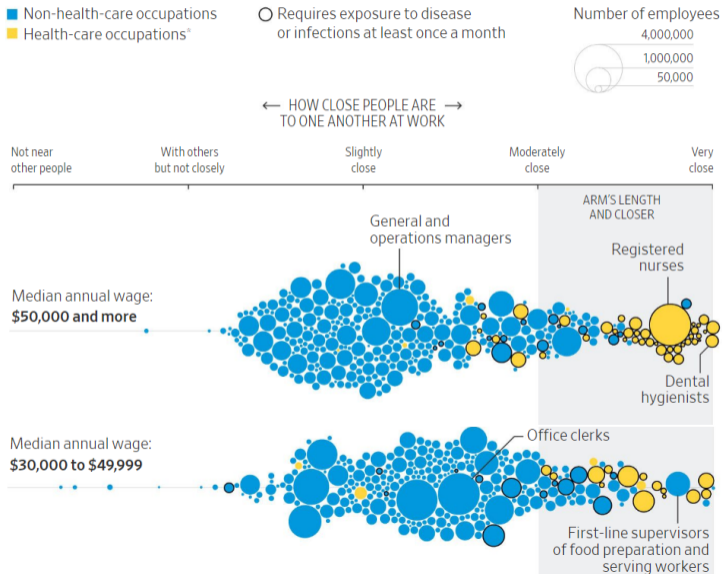
Pros

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

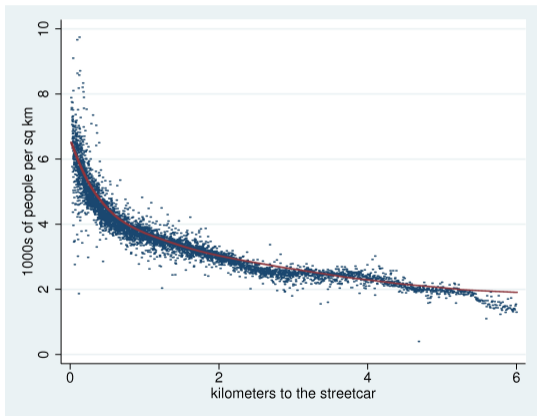
Cons

- 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



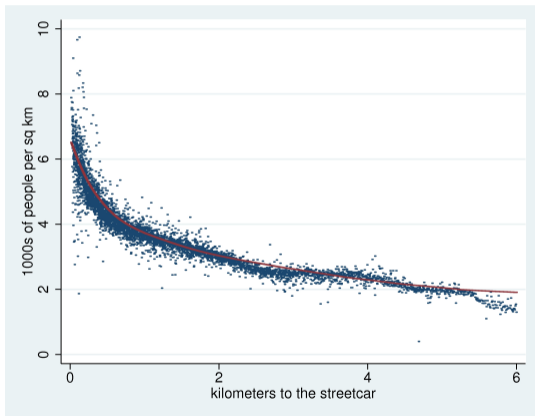
My Best Ever Scatter



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

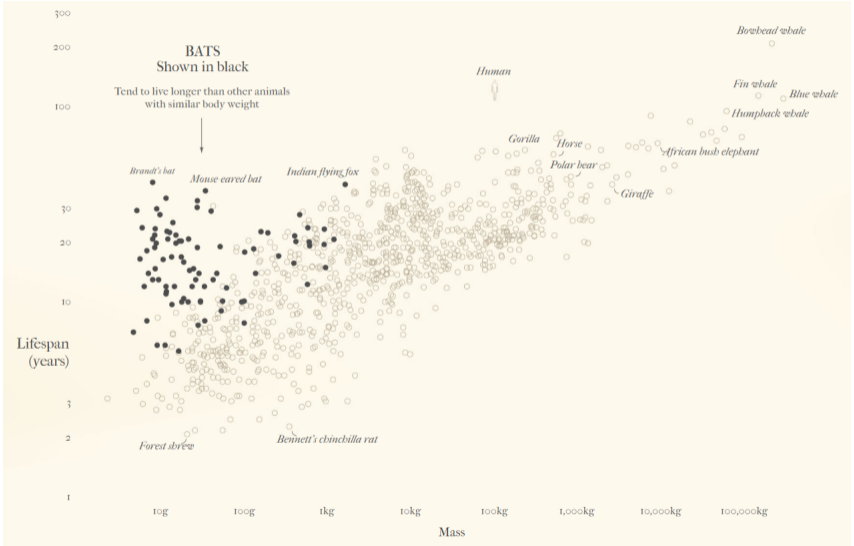


What is it?

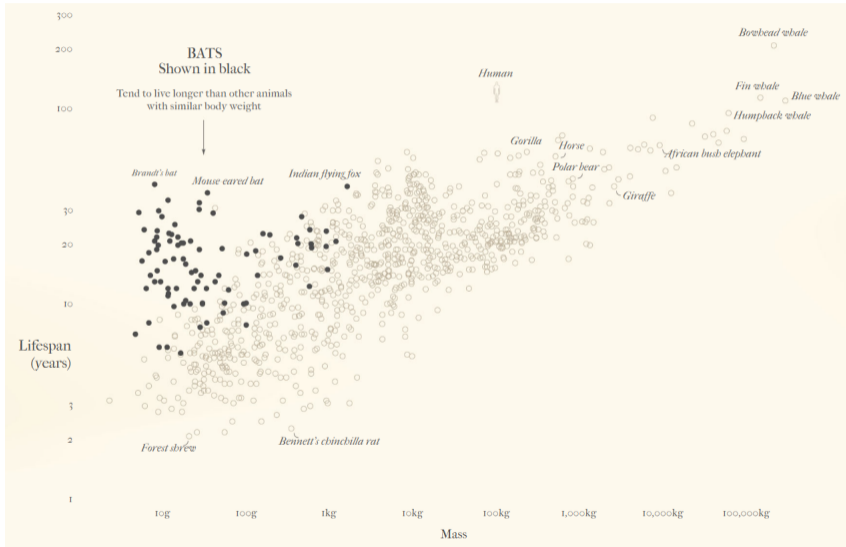
- Each point is
- average population density near about 400 land plots
- at a given distance from an old streetcar
- red line is a flexible regression line

Data show the point

How Can You Make a Scatter Explicible?



How Can You Make a Scatter Explicible?

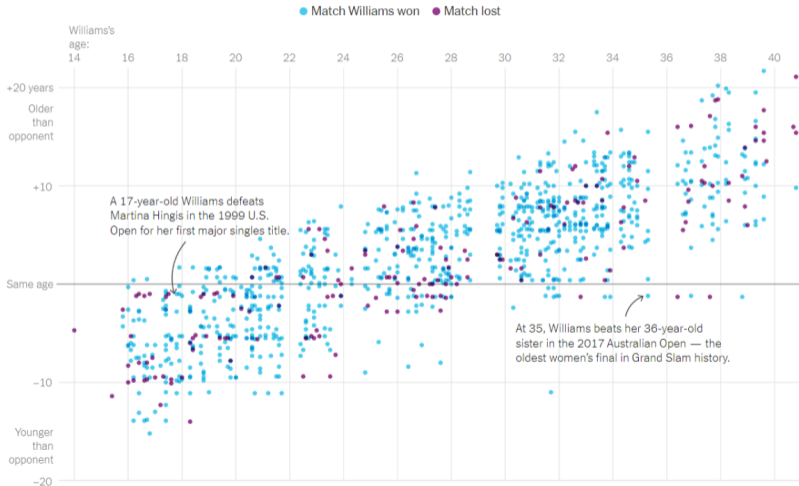


call out individual items

From Reuters

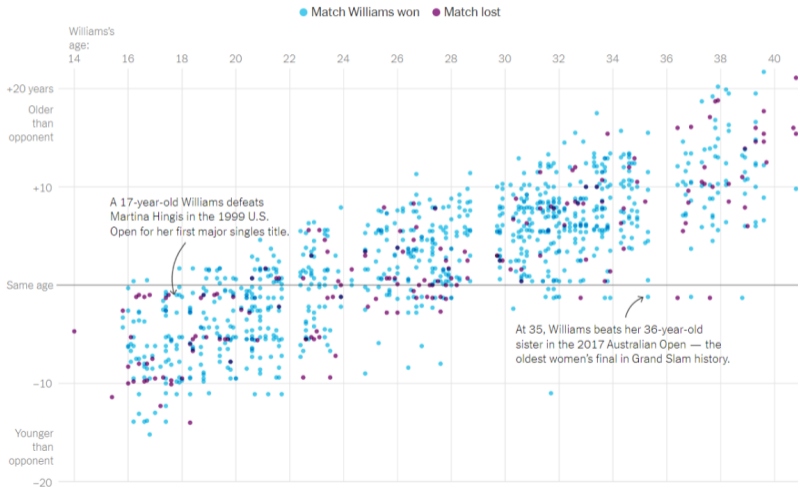
How Can You Make a Scatter Explicable?

Serena Williams's Professional Matches, Age 14 to 40



How Can You Make a Scatter Explicable?

Serena Williams's Professional Matches, Age 14 to 40



colors & call out individual items

From the *New York Times*

How Can You Make a Scatter Explicable?

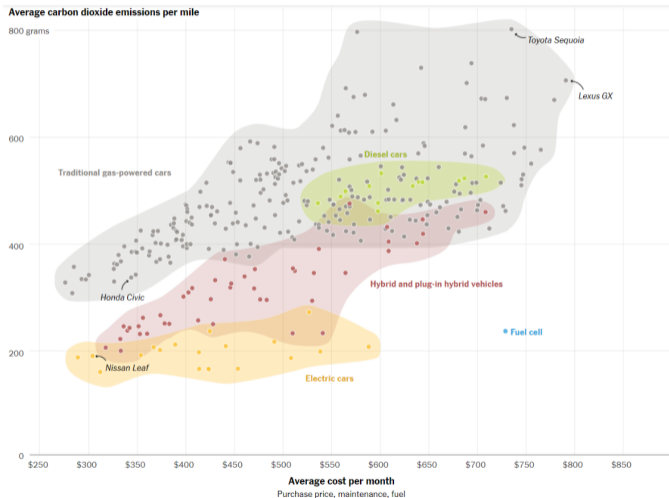
Covid vaccination rate by country



mark axes

From the *New York Times*

How Can You Make a Scatter Explicable?



circle common
points

From the [New York Times](#)

Going From Too Confusing Scatter to Bar Chart

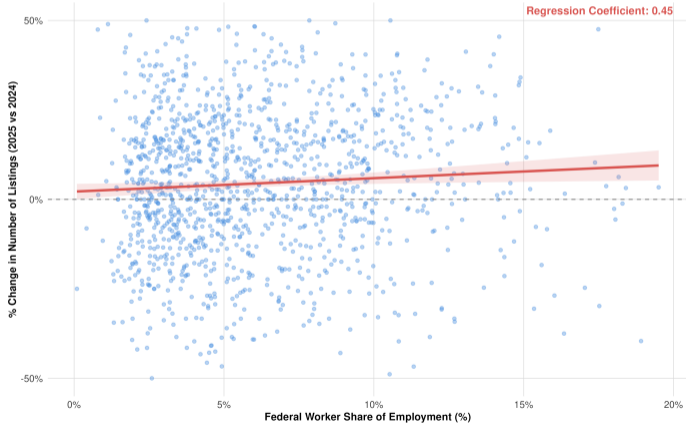
- Goal: Do neighborhoods with more federal workers have greater declines in home prices after Trump arrival?
- First pass: Scatter neighborhood share of
- After many tries: Bar, but still not great

First Scatter Version

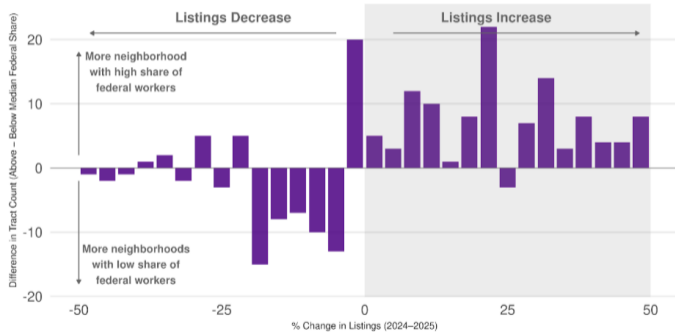
Change in Home Listings (2025 vs 2024) by Federal Worker Share

Year-to-date through October 17 | LODES Federal Share (2022)

Data Sources: Bright MLS listings, LODES 2022 RAC data by block group



Millionth Bar Version



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

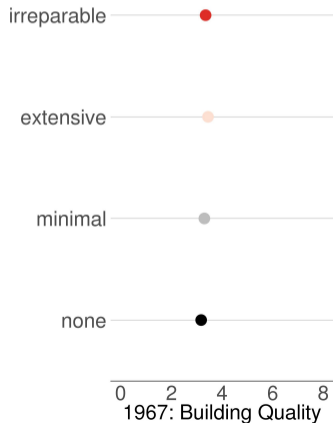
When do you use them?

- 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.

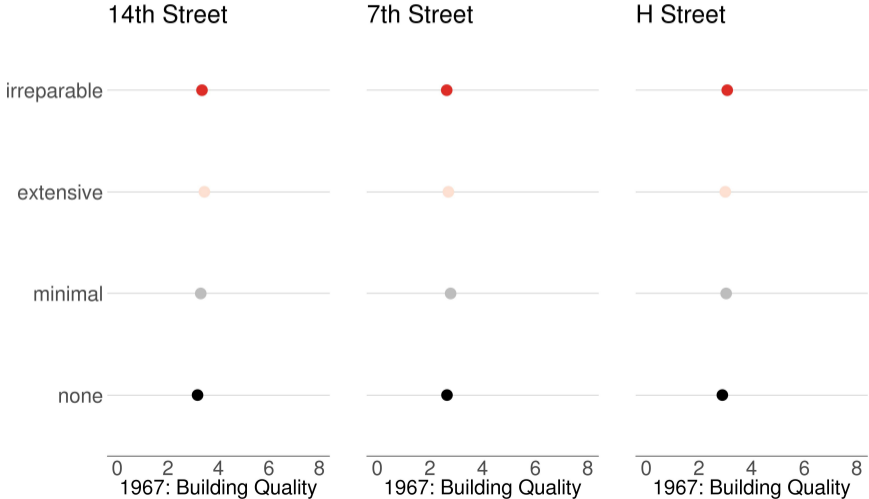
Destruction Roughly Even by 1967 Quality

14th Street



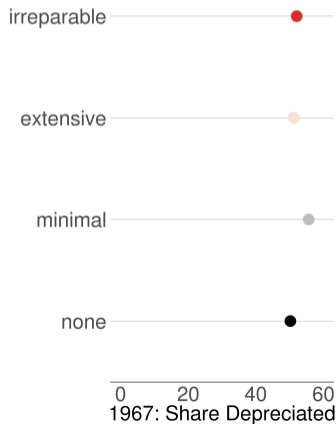
My Small Multiples

Destruction Roughly Even by 1967 Quality



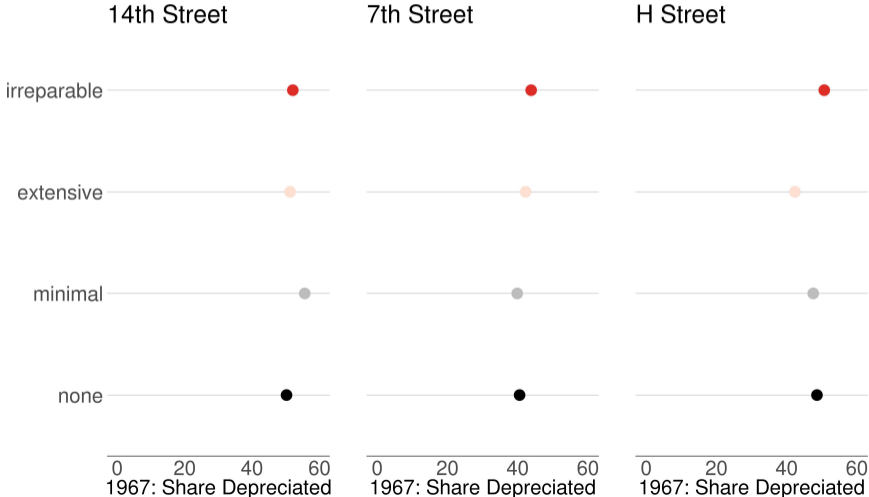
Destruction Roughly Even by 1967 Depreciation

14th Street



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 → 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 \rightarrow darker = larger \rightarrow 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, met Brewer

With thanks to [Cynthia Brewer](#), [Towards Data Science](#), datawrapperr.de, and [this Adobe blog](#).

Axes, Gridlines, Tickmarks, Axis Lines, Borders

All Other Little Graph Bits

ggplot2 Theme Elements

theme(element_name = element_function())

- element_text()
- element_line()
- element_rect()
- element_blank()

Plot elements:

plot.background
element_rect()

plot.title
element_text()

plot.margin
margin()

Facetting elements:

strip.background
element_rect()

panel.spacing
unit()

strip.text
element_text()

Axis elements:

axis.ticks
element_line()

axis.title
element_text()

axis.text
element_text()

axis.line
element_line()

Legend elements:

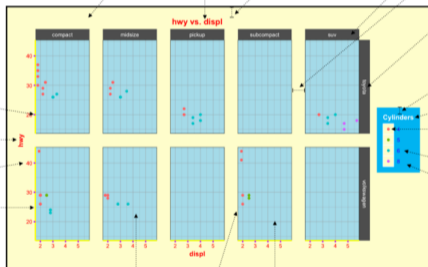
legend.margin
margin()

legend.title
element_text()

legend.key
element_rect()

legend.text
element_text()

legend.background
element_rect()



panel.background
element_rect()

panel.border
element_rect(fill = NA)

panel.grid
element_line()

Panel elements:

henrywang.nl

Derived from "ggplot2: Elegant Graphics for Data Analysis"

Rules of thumb

- Omit what you can
- Use grey when possible

Thank you, [Henry Wang](#)

R Notes on Scatters

Scatter Recap

- Scatters show correlation between two variables
- Powerful
- Hard for laypeople to interpret – so you should
 - heavily annotate OR
 - convert to more comprehensible graphic

Next Lectures

- Guest speaker and workshop
- Interactivity, and then presentations
- Presentations due by noon day of presentation
- Final paper due May 4 by midnight