

Lecture 11: Scatter Plots

April 22, 2019

Overview

Course Administration

Good, Bad and Ugly

Scatter Plots

Small Multiples, or Facets

R Notes

Course Administration

1. Rest of the class, next week
 - Monday: final tutorial, due May ?
 - Wednesday: presentations 1
 - Thursday: presentations 2, room TBA
2. Presentation dates are assigned: if your group is not grouped, let me know
3. Will try to have all assignment grades for you to check by next Monday on Jill's sheet
4. Paper due May 3 by 5 pm
 - soft copy due to google drive
 - hard copy to due my mailbox by 9:15 AM Wed. May 8
5. Anything else?

This Week's Good Bad and Ugly

- ME
- ID

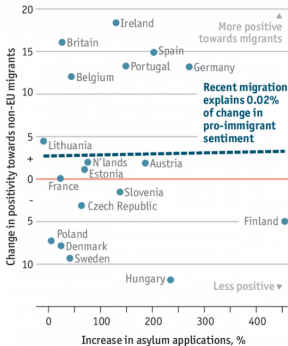
Marissa's Example

Unfamiliarity breeds contempt

Sentiment towards non-EU migrants*, 2014-18, percentage-point change

Increase in asylum applications

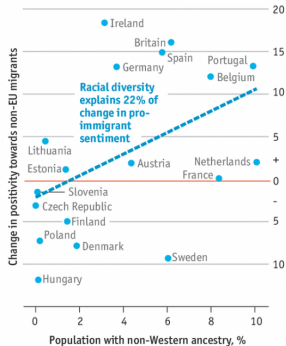
2013-14 to 2015-16



Sources: Eurobarometer; Eurostat; European Social Survey

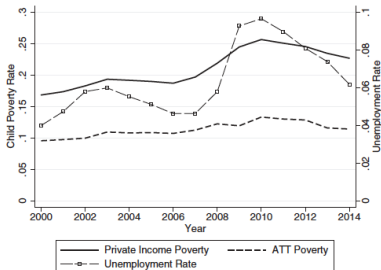
Share of population with non-Western ancestry

2014, %



*Selected EU countries where ancestry data is available

Ilhams's Example



From Bitler, Marianne, Hilary Hoynes, and Elira Kuka. Child Poverty, the Great Recession, and the Social Safety Net in the United States. *Journal of Policy Analysis and Management* 36, no. 2 (2017): 358-89.

Scatter Plots

What Does a Scatter Do?

What Does a Scatter Do?

- Shows correlation between two items

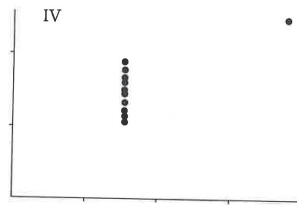
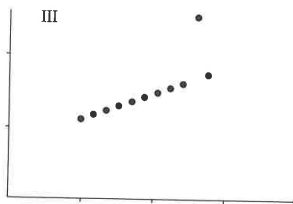
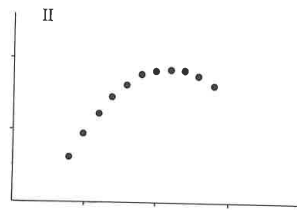
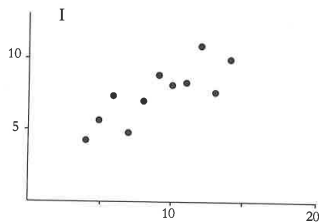
What Does a Scatter Do?

- Shows correlation between two items
- Most common type of graph for academic presentation
- Requires the audience to think about the relationship
- Not always desirable for policy communication

A Reminder: Anscombe's Quartet

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: Anscombe's Quartet



When Did They First Appear?

- Some early work by physicists in the 1700s
- As we know them now between 1906 and 1920 (Friendly and Denis, 2005)

A Very Early Scatter

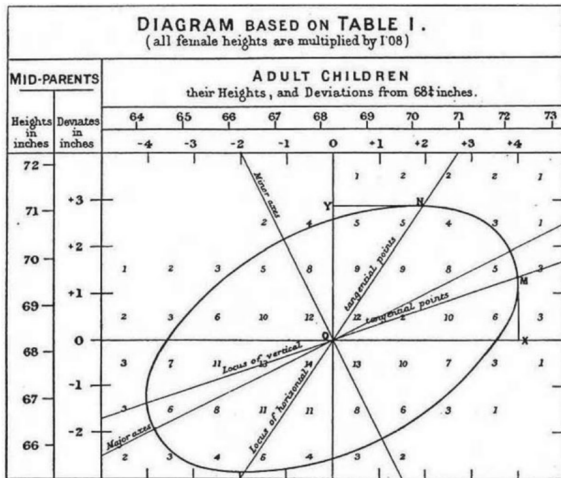


FIGURE 7.

Galton's smoothed correlation diagram for the data on heights of parents and children, showing one ellipse of equal frequency. Source: Galton (1886, Plate X).

How Can You Annotate a Scatter?

How Can You Annotate a Scatter?

- best fit lines
- ovals
- colors

Small Multiples, or Facets

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
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?
 - 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?

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.

How Beyonce Exploits the Power of Small Multiples

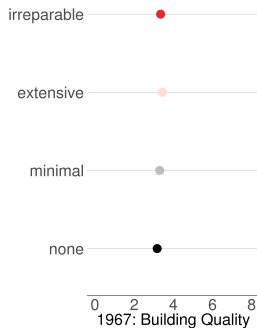


With thanks to [Vibe](#).

My Small Multiples

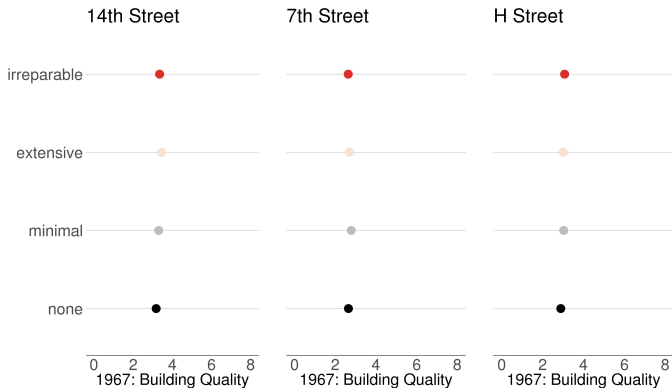
Destruction Roughly Even by 1967 Quality

14th Street



My Small Multiples

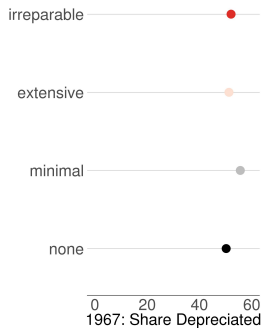
Destruction Roughly Even by 1967 Quality



My Small Multiples

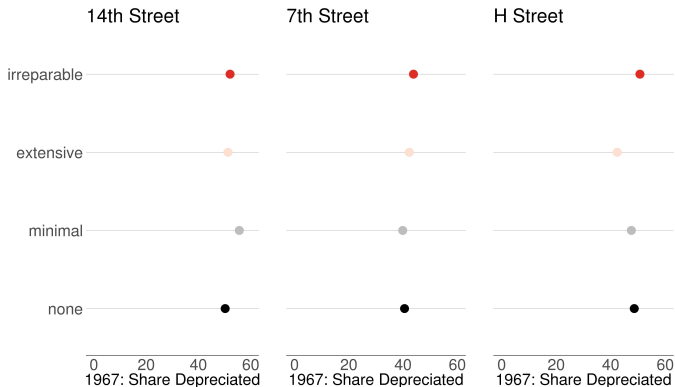
Destruction Roughly Even by 1967 Depreciation

14th Street



My Small Multiples

Destruction Roughly Even by 1967 Depreciation



R Notes

Today in R: Line Charts and De-Bugging

1. Scatter plots: `geom_point()`
2. Segments: `geom_segment()`
3. Small multiples
4. Instead of a loop: Use vector power

1. Scatter plots

```
p1 <- ggplot() +  
  geom_point(data = df,  
            mapping = aes(x = xvar, y = yvar))
```

Scatter plots: Shapes

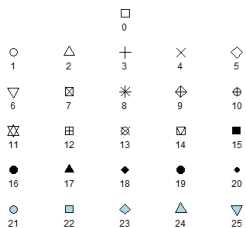


Figure 1:

Scatter plots: Shapes

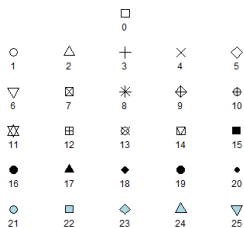


Figure 1:

```
p1 <- ggplot() +  
  geom_point(data = df,  
            mapping = aes(x = xvar, y = yvar),  
            shape = SHAPE.NUMBER)
```

Scatter plots: One color

```
p1 <- ggplot() +  
  geom_line(data = polys,  
            mapping = aes(x = xvar, y = yvar),  
            color = "COLOR.NAME")
```


Scatter plots: Colors by Group

```
p1 <- ggplot() +  
  geom_line(data = polys,  
            mapping = aes(x = xvar, y = yvar),  
            color = VARIABLE)
```

Scatter plots: Colors by Group

```
p1 <- ggplot() +  
  geom_line(data = polys,  
            mapping = aes(x = xvar, y = yvar),  
            color = VARIABLE)
```

- ▶ To show colors by a variable
- ▶ You can specify colors in

```
scale_colour_manual(values=c('A'='grey',  
                             'E'='red',  
                             'F'='blue'))
```

Scatter plots: Calling out Regions

- ▶ best fit line: use cautiously
`geom_smooth(method = lm, se = FALSE)`
- ▶ best fit curve: same
`geom_smooth(se = FALSE)`
- ▶ best fit curve: with shaded error region
`geom_smooth()`
- ▶ annotations
`geom_rect()` `geom_segment()`

2. Drawing Segments

This is a scatterplot with segments!

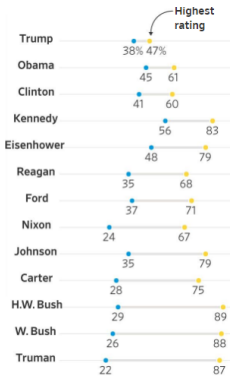


Figure 2:

Thanks to [WSJ](#).

Code Segments

```
s2 <- ggplot() +  
  geom_segment(data = df,  
              mapping = aes(x = VARIABLE1,  
                            xend = VARIABLE2,  
                            y = VARIABLE3,  
                            yend = VARIABLE4))
```

Code Segments

```
s2 <- ggplot() +  
  geom_segment(data = df,  
              mapping = aes(x = VARIABLE1,  
                            xend = VARIABLE2,  
                            y = VARIABLE3,  
                            yend = VARIABLE4))
```

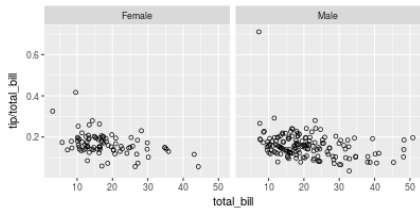
There is also `geom_curve` for brave people

3. Small Multiples, or Facets

```
facet_grid(rows = vars(VARIABLE))
```

3. Small Multiples, or Facets

```
facet_grid(rows = vars(VARIABLE))
```



Thanks to [Winston Chang](#).

Facet Columns

```
facet_grid(cols = vars(VARIABLE))
```

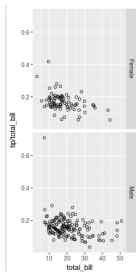


Figure 3:

Or both.

4. Avoiding a Loop

Suppose you want to do this many times

```
df$ln.x <- log(df$x)
```

4. Avoiding a Loop

Suppose you want to do this many times

```
df$ln.x <- log(df$x)
```

This does not work!

```
tolog <- c(x,y,z)
for(i in tolog){
  df$ln.i <- log(df$i)
}
```

and you can't fix it up with `eval(parse())` either.

The Elegant Solution

```
tolog <- c("x","y","z")  
df[paste0("ln.",tolog)] <- log(df[tolog])
```

The Elegant Solution

```
tolog <- c("x", "y", "z")  
df[paste0("ln.", tolog)] <- log(df[tolog])
```

Recall:

$$y = \log_b(x)$$

and

$$x = b^y$$

The Elegant Solution in Action

```
df <- data.frame(x = c(1, 2, 3),  
                 y = c(10, 20, 30),  
                 z = c(100, 200, 300))
```

The Elegant Solution in Action

```
df <- data.frame(x = c(1, 2, 3),  
                 y = c(10, 20, 30),  
                 z = c(100, 200, 300))
```

```
df
```

```
##   x  y  z  
## 1 1 10 100  
## 2 2 20 200  
## 3 3 30 300
```

The Elegant Solution in Action

```
df <- data.frame(x = c(1, 2, 3),
                 y = c(10, 20, 30),
                 z = c(100, 200, 300))
tolog <- c("x", "y", "z")
df[paste0("ln.", tolog)] <- log(df[tolog])
df
```

```
##   x  y  z      ln.x      ln.y      ln.z
## 1 1 10 100 0.0000000 2.302585 4.605170
## 2 2 20 200 0.6931472 2.995732 5.298317
## 3 3 30 300 1.0986123 3.401197 5.703782
```


Next Lectures

- Check presentation dates for group togetherness
- Policy brief due May 3 at 5 pm