WELCOME!

GETTING STARTED

Open an issue in your Assignments Repository letting me know how far you got with PS-07 and what issues (if any) you had!

Check out the new website at https://slu-soc5050.github.io

Check GitHub for assignment feedback, midterm grade reports

FALL, 2017

CHRISTOPHER PRENER, PH.D.

LECTURE 11

WEEK 10

QUANTITATIVE ANALYSIS

CORRELATION (PART 1)

AGENDA

1. Front Matter
2. Public Polling
3. Handouts
4. Covariance
5. Scatterplots
6. Pearson's *r*
7. Back Matter

1 FRONT MATTER

ANNOUNCEMENTS

Jotter/wiki has been depreciated - all posts have been migrated to the new website and new content will only be posted there!

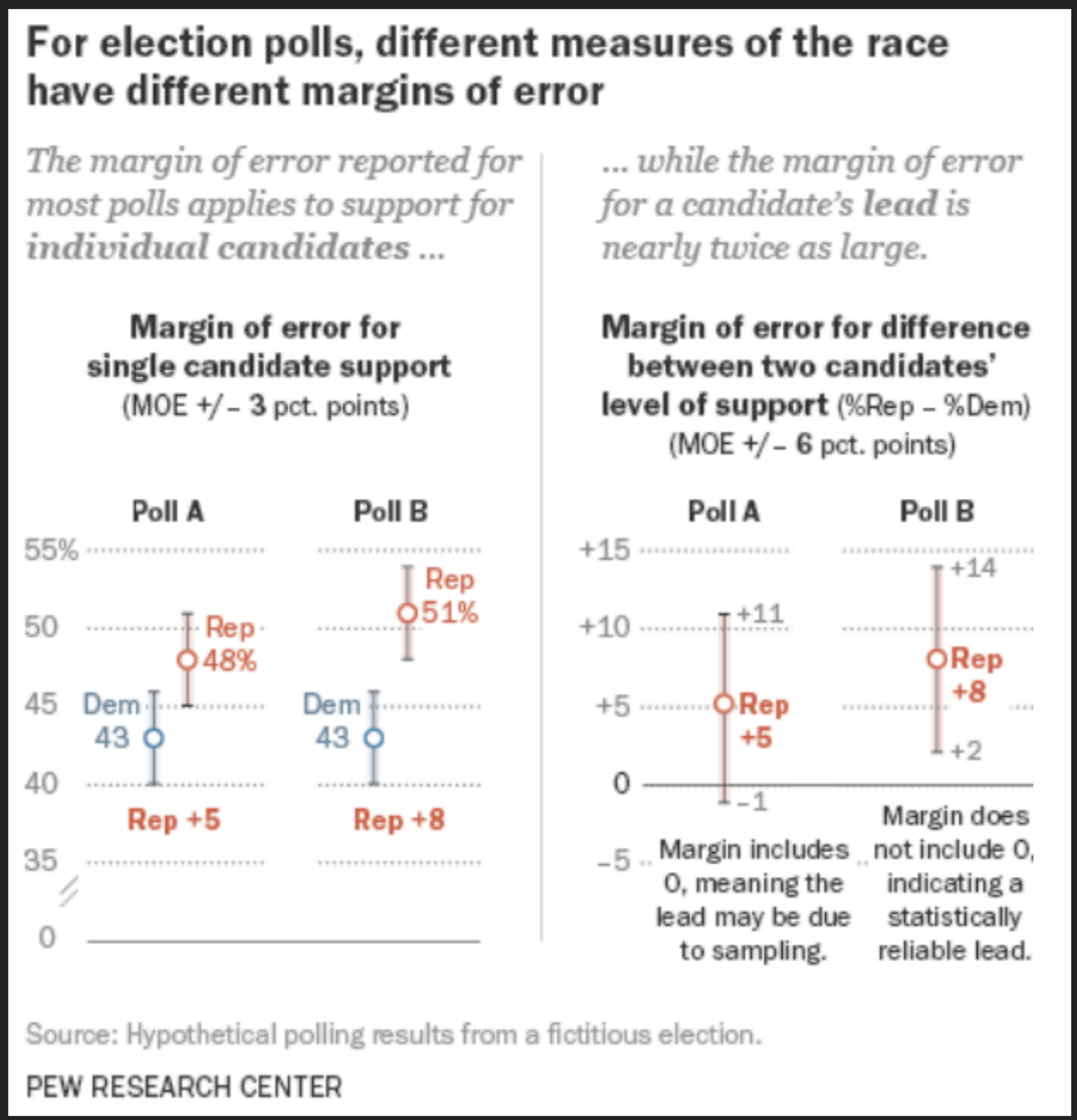
Lab-10 due next Monday

Lab-11 and PS-08 due Monday, 11/13

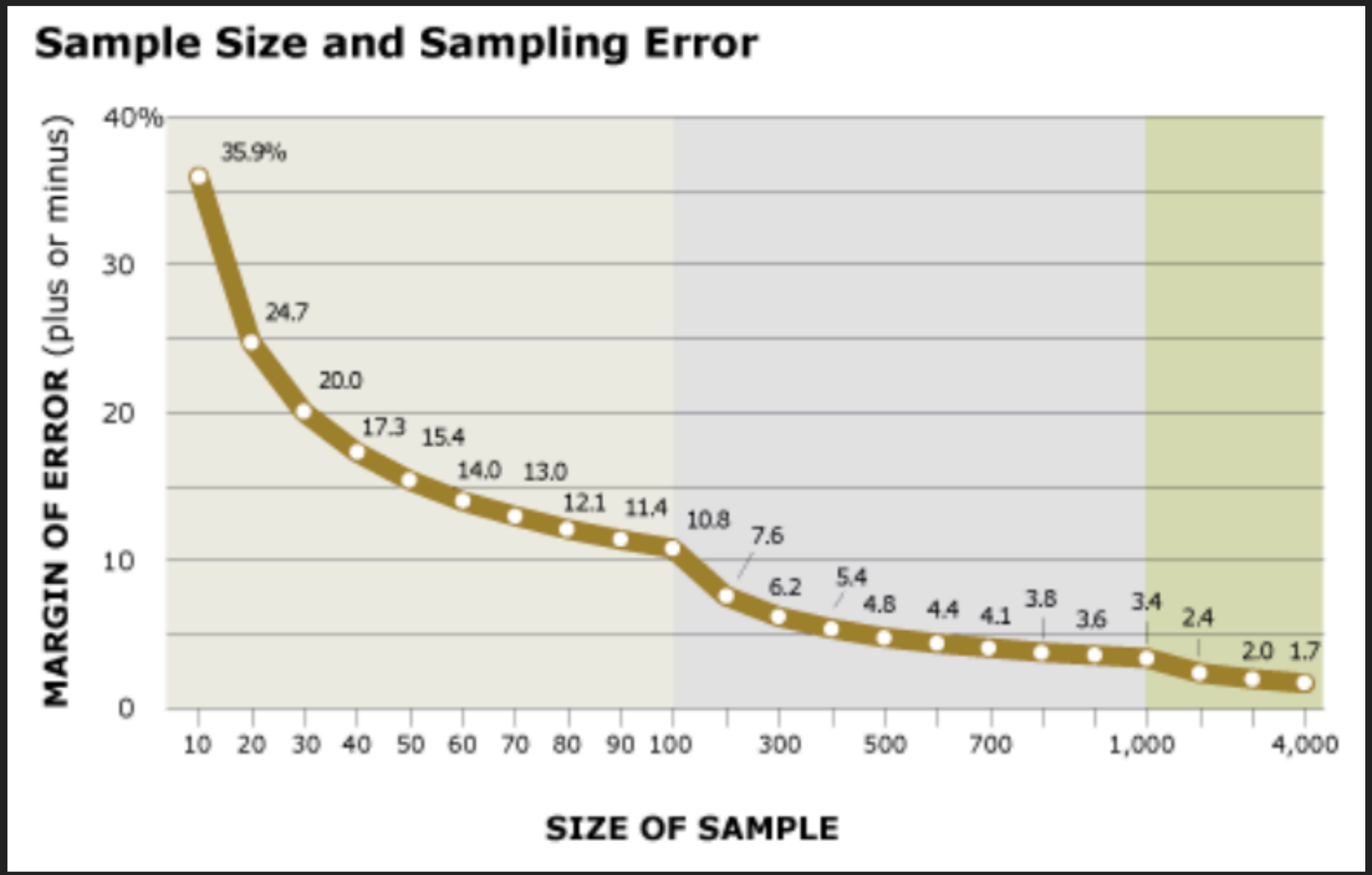
Final project drafts due Monday, 11/13 (everyone should have a handout and draft slides; SOC 5050 students should also have a draft paper to submit)

2 PUBLIC POLLING

INTERPRETING A GIVEN POLL



MARGIN OF ERROR AND N



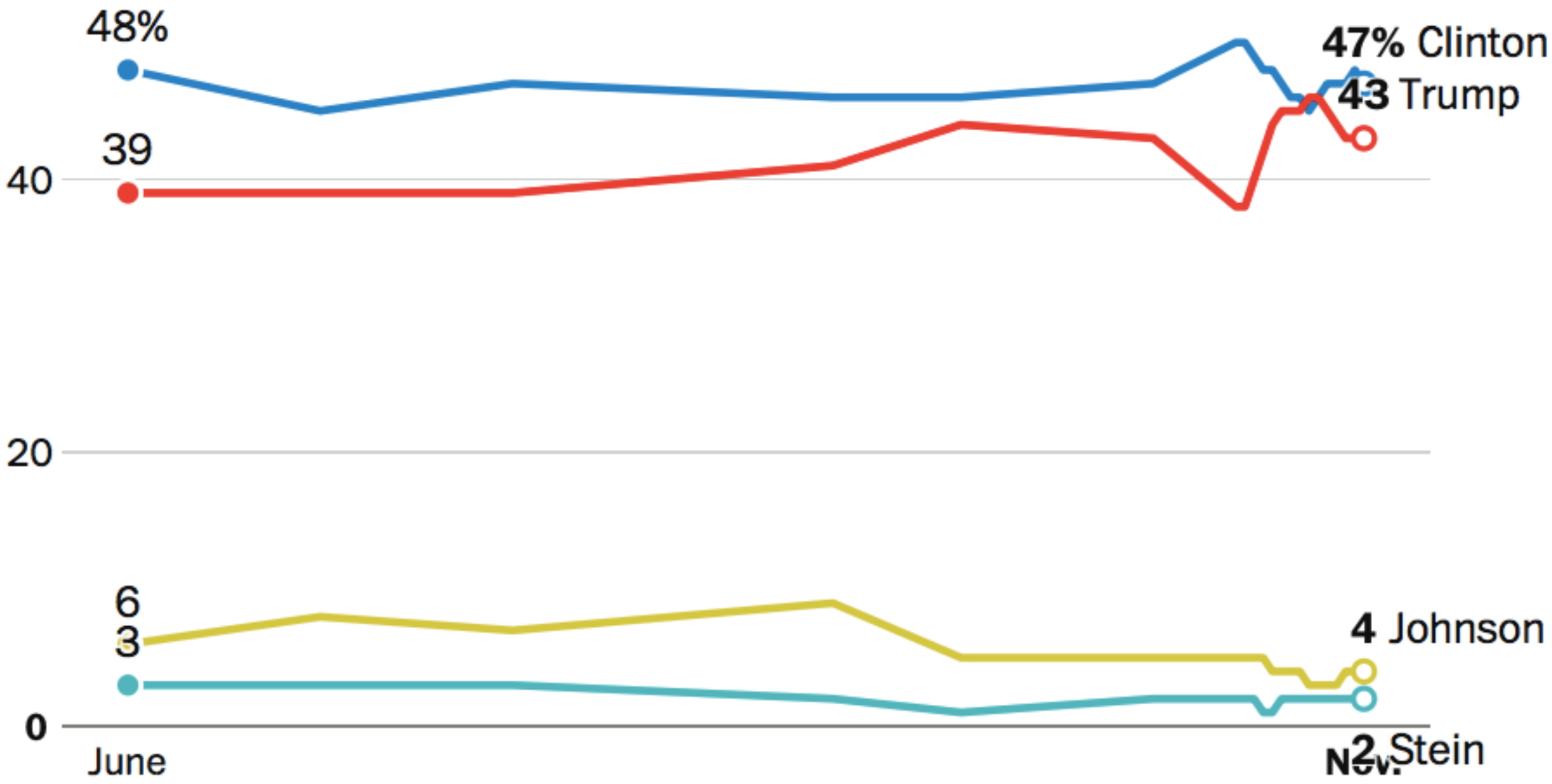
INTERPRETING A GIVEN POLL

Post-ABC Tracking Poll, Nov. 2-5, 2016

This Washington Post-ABC News poll was conducted by telephone November 2-5, 2016, among a random national sample of 2,854 adults including landline and cell phone respondents. Overall results have a margin of sampling error of plus or minus two points; the error margin is plus or minus 2.5 points among the sample of 1,937 likely voters. Sampling, data collection and tabulation by Abt-SRBI of New York.

Clinton's lead stable for past three days of Post-ABC Tracking Poll

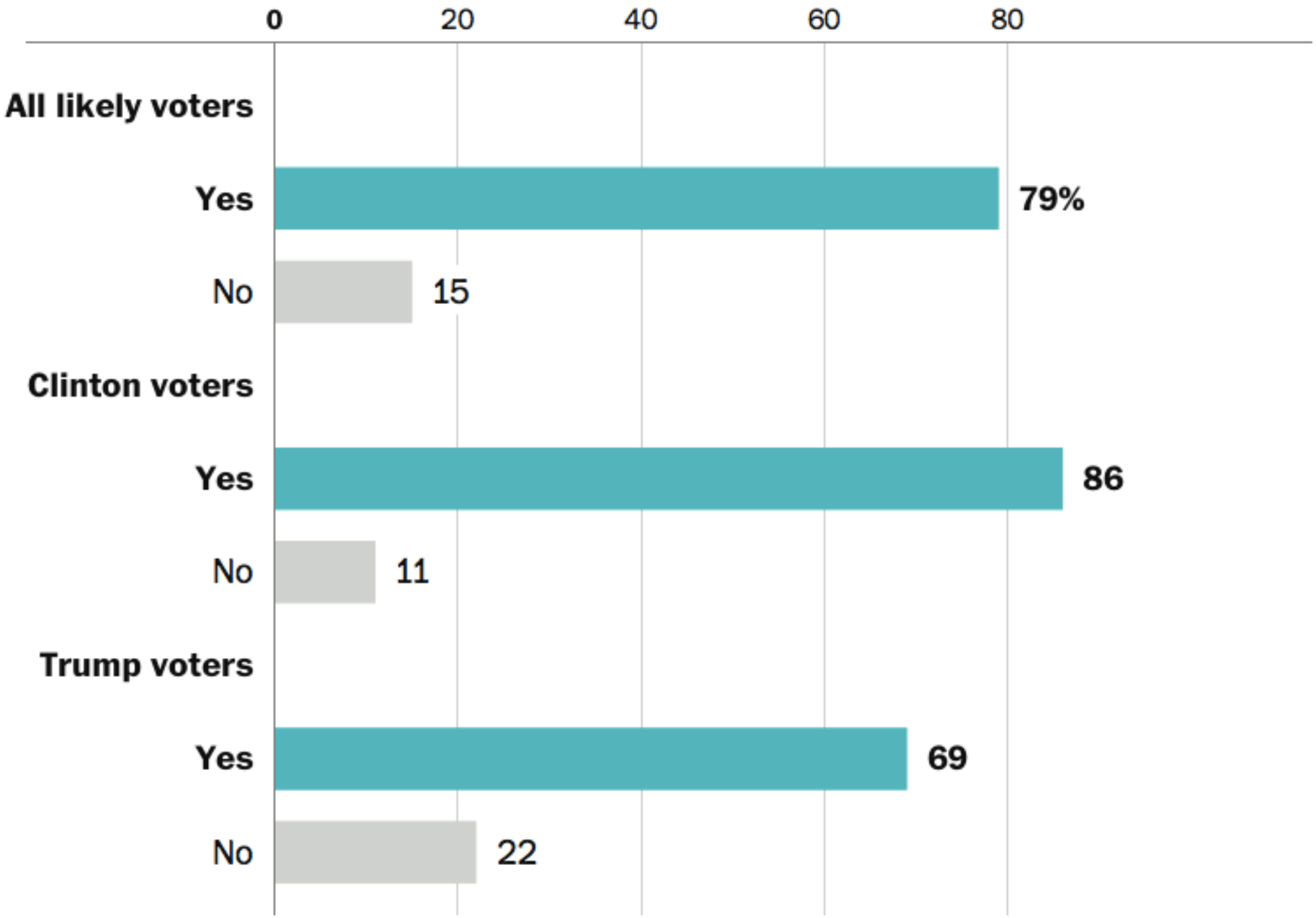
Among likely voters



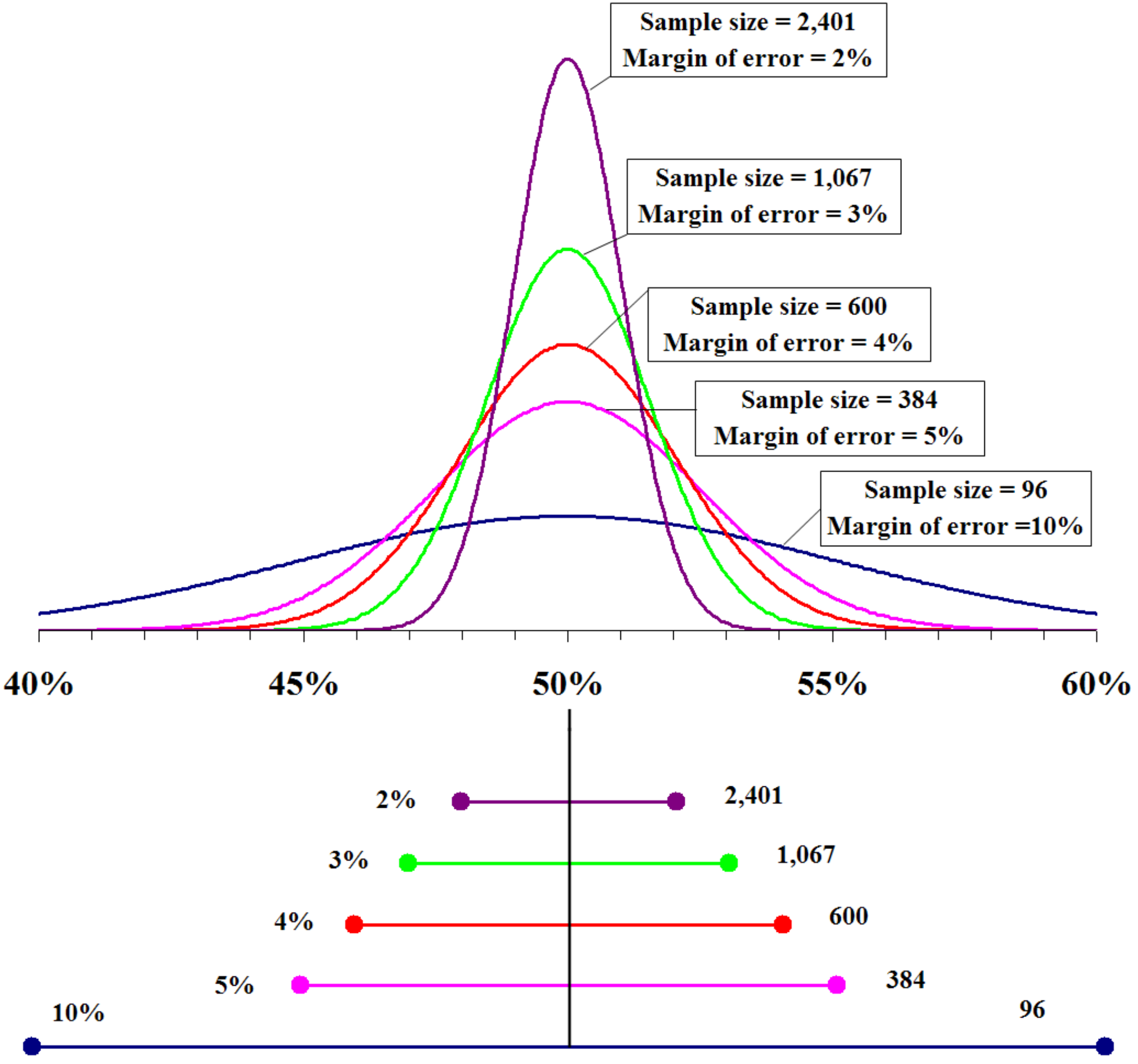
Source: Post-ABC polls June-November

Trump supporters far less likely to say they will accept the results of the election than Clinton supporters

Q: Regardless of whom you support, are you prepared to accept the outcome of the election as legitimate, or are you not prepared to do that at this time? (Among likely voters)



Source: Washington Post-ABC News Tracking Poll Nov. 3-5 among 1,763 likely voters; margin of error +/- 2.5 points

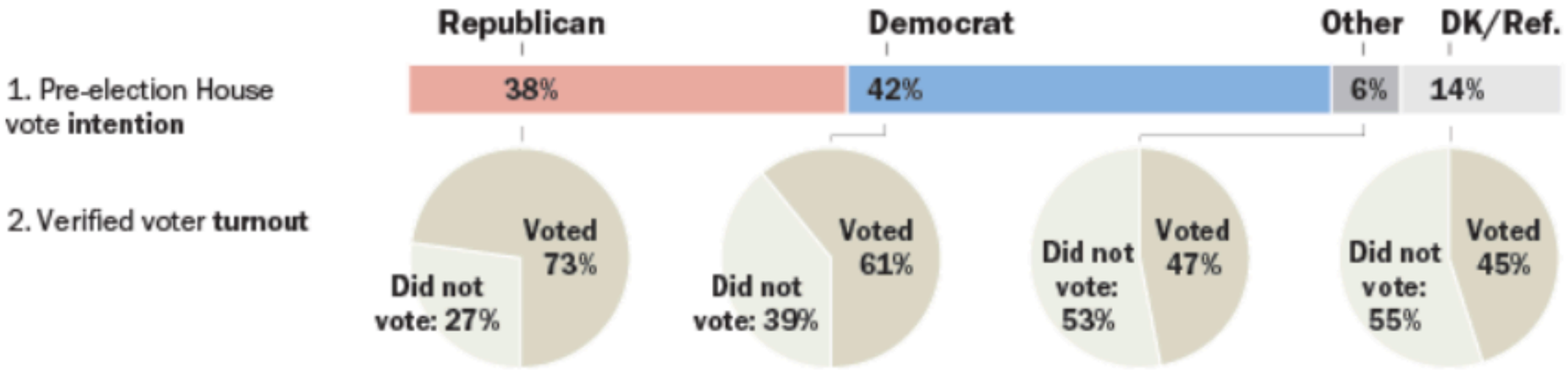

Sample size = 2,401
Margin of error = 2%
Sample size = 1,067
Margin of error = 3%
Sample size = 600
Margin of error = 4%
Sample size = 384
Margin of error = 5%
Sample size = 96
Margin of error =10%
40%
45%
50%
55%
60%
2%
2,401
3%
1,067
4%
600
5%
384
10%
96

SOURCES OF ERROR: INTENTION VS. TURNOUT

Vote intention and turnout

Vote intention and turnout

Republican candidates benefited from higher turnout by their supporters. Among pre-election respondents who supported Republican candidates, 73% turned out to vote; among those who supported Democratic candidates, 61% turned out to vote.



Source: 2014 American Trends Panel September and November waves. Based on registered voters who participated in both waves and were matched to a national voter file.

PEW RESEARCH CENTER

SOURCES OF ERROR: LIKELY VOTER SCREENS

- How much thought have you given to the coming November election? ***Quite a lot, some,*** *only a little, none*
- Have you ever voted in your precinct or election district? ***Yes,*** *no*
- Would you say you follow what's going on in government and public affairs ***most of the time, some of the time,*** *only now and then, hardly at all?*
- How often would you say you vote? ***Always, nearly always,*** *part of the time, seldom*
- How likely are you to vote in the general election this November? ***Definitely will vote, probably will vote,*** *probably will not vote, definitely will not vote*
- In the 2012 presidential election between Barack Obama and Mitt Romney, did things come up that kept you from voting, or did you happen to vote? ***Yes, voted;*** *no*
- Please rate your chance of voting in November on a scale of 10 to 1. *0-8,* ***9, 10***

SOURCES OF ERROR: LIKELY VOTER SCREENS

The Perry-Gallup index

Score on scale	Share of registered voters	Share of all verified voters	% who are verified voters in each group
7	48%	63%	83%
6	15	15	63
5	10	10	59
4	7	4	34
3	6	4	41
2	6	2	23
1	3	1	13
0	4	1	11
	100	100	

83% of those who scored a 7 actually voted

22% of verified voters scored between 0 and 5

Source: 2014 American Trends Panel September and November waves. Based on registered voters who participated in both waves and were matched to a national voter.

PEW RESEARCH CENTER

SOURCES OF ERROR: LIKELY VOTER SCREENS

The Upshot | The New York Times

We Gave Four Good Pollsters the Same Raw Data. They Had Four Different Results.

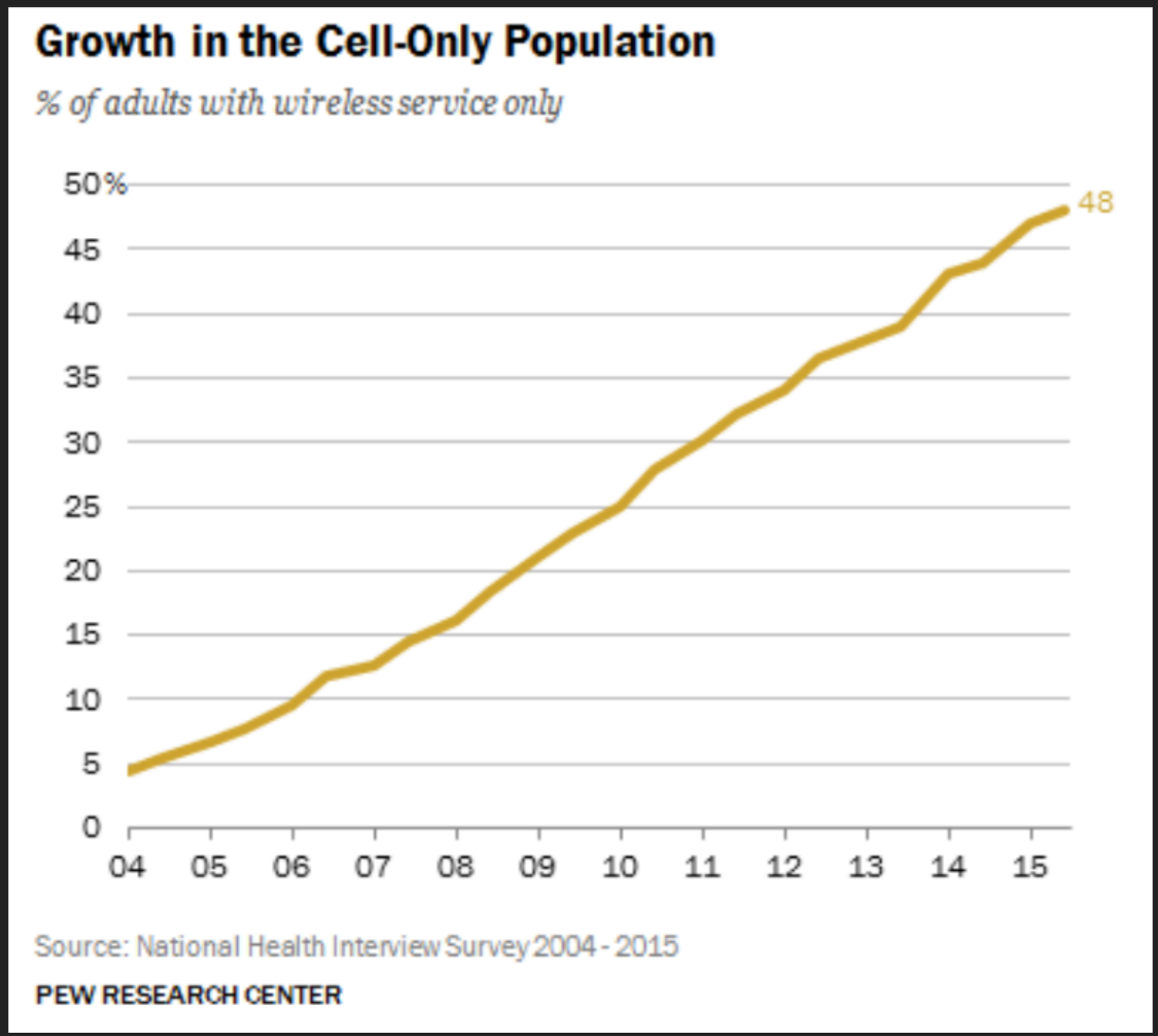
By **NATE COHN** SEPT. 20, 2016

How four pollsters, and The Upshot, interpreted 867 poll responses:

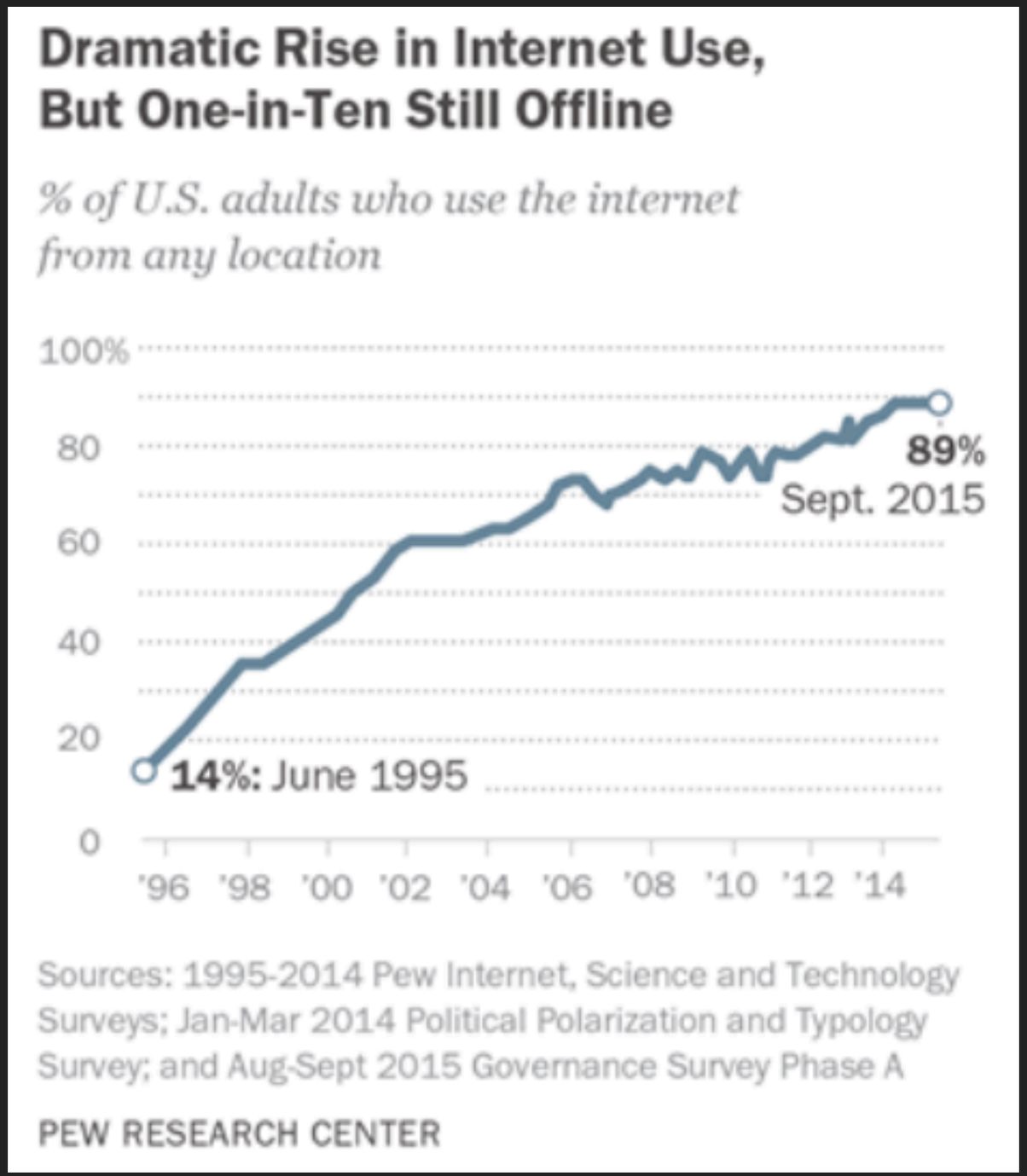
SOURCES OF ERROR: LIKELY VOTER SCREENS

Pollster	Who is Likely Voter?	Type of weight	Tries to match...
Charles Franklin Marquette Law	Self-report	Traditional	Census
Patrick Ruffini Echelon Insights	Vote history	Traditional	Voter File
Omero, Green, Rosenblatt Penn Schoen Berland Research	Self-report	Traditional	Voter File
Corbett-Davies, Gelman, Rothschild Stanford University/Columbia University/Microsoft Research	Vote history	Model	Voter File
NYT Upshot/Siena College	Report + history	Traditional	Voter File

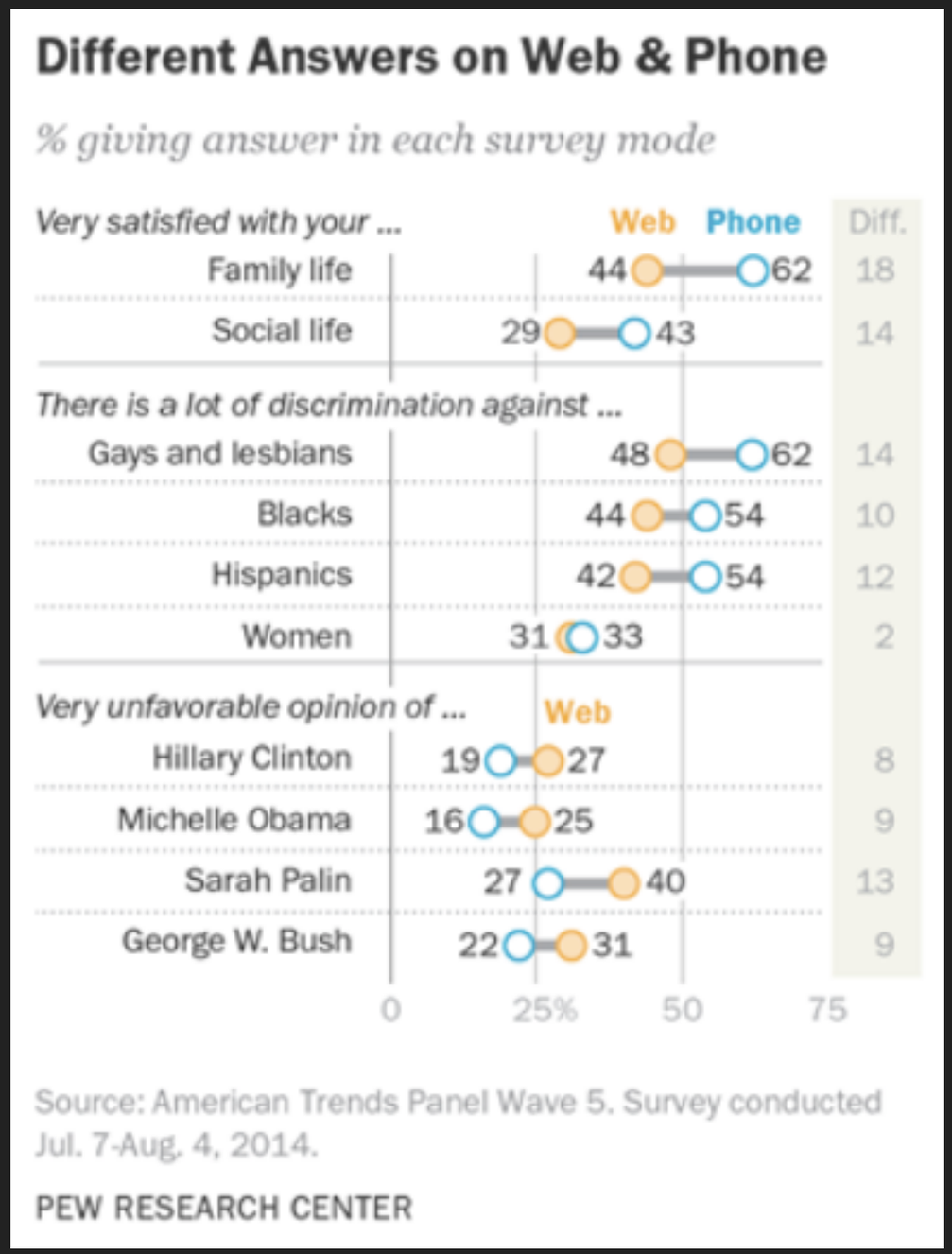
SOURCES OF ERROR: SURVEY MECHANISM



SOURCES OF ERROR: SURVEY MECHANISM



SOURCES OF ERROR: SOCIAL DESIRABILITY



SOURCES OF ERROR: NON-RESPONSE

Surveys Face Growing Difficulty Reaching, Persuading Potential Respondents

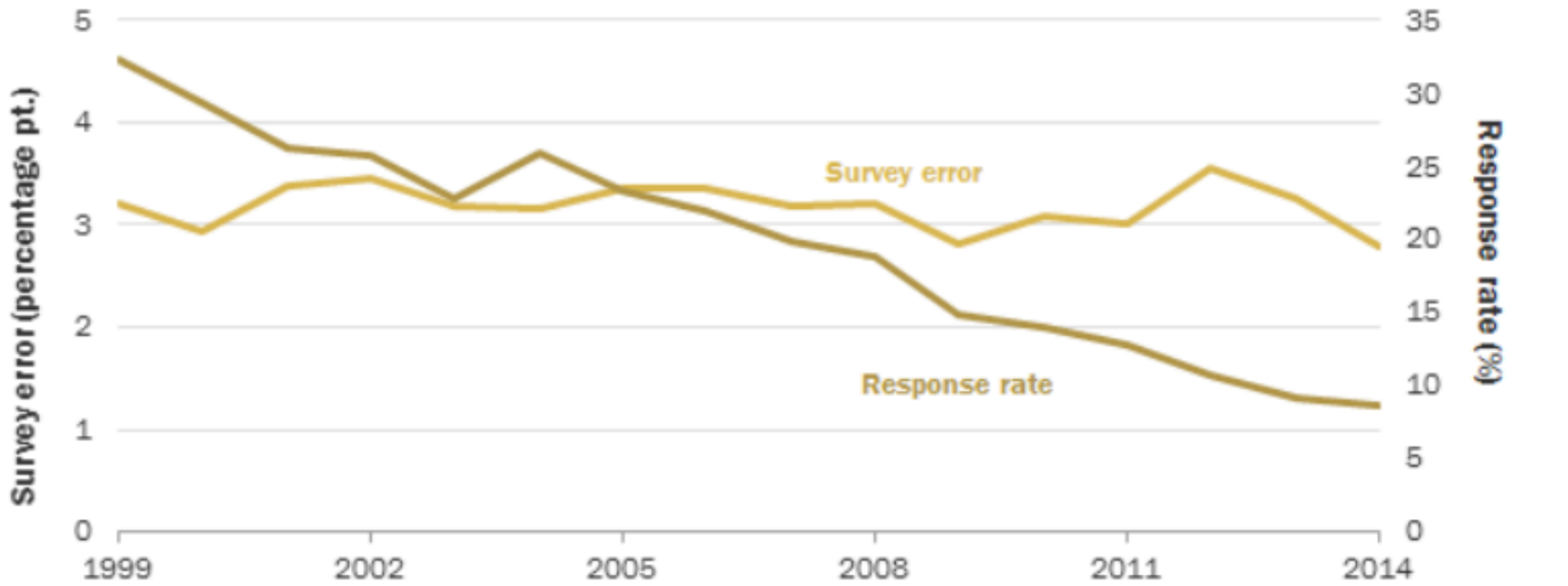
	1997 %	2000 %	2003 %	2006 %	2009 %	2012 %
Contact rate (percent of households in which an adult was reached)	90	77	79	73	72	62
Cooperation rate (percent of households contacted that yielded an interview)	43	40	34	31	21	14
Response rate (percent of households sampled that yielded an interview)	36	28	25	21	15	9

PEW RESEARCH CENTER 2012 Methodology Study. Rates computed according to American Association for Public Opinion Research (AAPOR) standard definitions for CON2, COOP3 and RR3. Rates are typical for surveys conducted in each year.

SOURCES OF ERROR: NON-RESPONSE

Despite falling response rates, error of telephone RDD polls unchanged since 1999

Percentage-point error in weighted poll estimates for the size of two-way demographics (i.e. men ages 18-34) when compared with data from the U.S. Census Bureau



Note: Survey error reported is the mean absolute bias of the marginal distributions of four demographic variables (Education, Age, Region, and Race). Error is measured by averaging estimates of two-way demographics (i.e. women ages 18-34) when compared with the U.S. Census Bureau.

Source: David Dutwin analysis of 15 years of surveys conducted by Pew Research Center, ABC/Washington Post and CBS/The New York Times, available at www.washingtonpost.com/news/the-fix/wp/2016/01/28/reports-of-the-demise-of-polls-have-been-greatly-exaggerated/

"Flashpoints in Polling"

PEW RESEARCH CENTER

SOURCES OF ERROR: SURVEY MECHANISM

Clinton does better in a forecast based only on live-interview polls

	CLINTON'S POLLS-ONLY FORECAST PROJECTION		
POLL TYPE	CHANCE OF WINNING	POPULAR VOTE MARGIN	ELECTORAL VOTES
Live	86.0%	+7.1	354.1
Nonlive	70.5	+4.8	315.0

All numbers rounded. Numbers as of the morning of Aug. 30.

SOURCE: FIVETHIRTYEIGHT POLLS-ONLY FORECAST

SOURCES OF ERROR: SURVEY MECHANISM

Most new swing-state polls show Clinton ahead

Clinton's margin in swing-state polls since 8 p.m. on Tuesday

	STATE	LIVE-CALLER POLLS	AUTO AND ONLINE POLLS
Clinton "firewall" states	Maine		+12
	Minnesota		+10
	Virginia	-3, +5	+10
	Wisconsin	+6	+2
	Michigan	+19	+1, +1, +3
	N. H.		+14, +4
	Pennsylvania	+5, +4, +2, +4	+6, +1
	Colorado		+3, +3, +7, +10
Other competitive states	Nevada	-6	0, -1, +7, 0
	N. C.	+3	+7, +3
	Florida	+1, +2	-4, +8, +3, 0
	Ohio	-5	-4, +3
	Iowa		-6
	Arizona	-5	-4, -1, -1
	Georgia		-9, -1
	Utah		-4

FIVETHIRTYEIGHT

SOURCES OF ERROR: DEMOGRAPHIC WEIGHTS

The Upshot | The New York Times

We Gave Four Good Pollsters the Same Raw Data. They Had Four Different Results.

By **NATE COHN** SEPT. 20, 2016

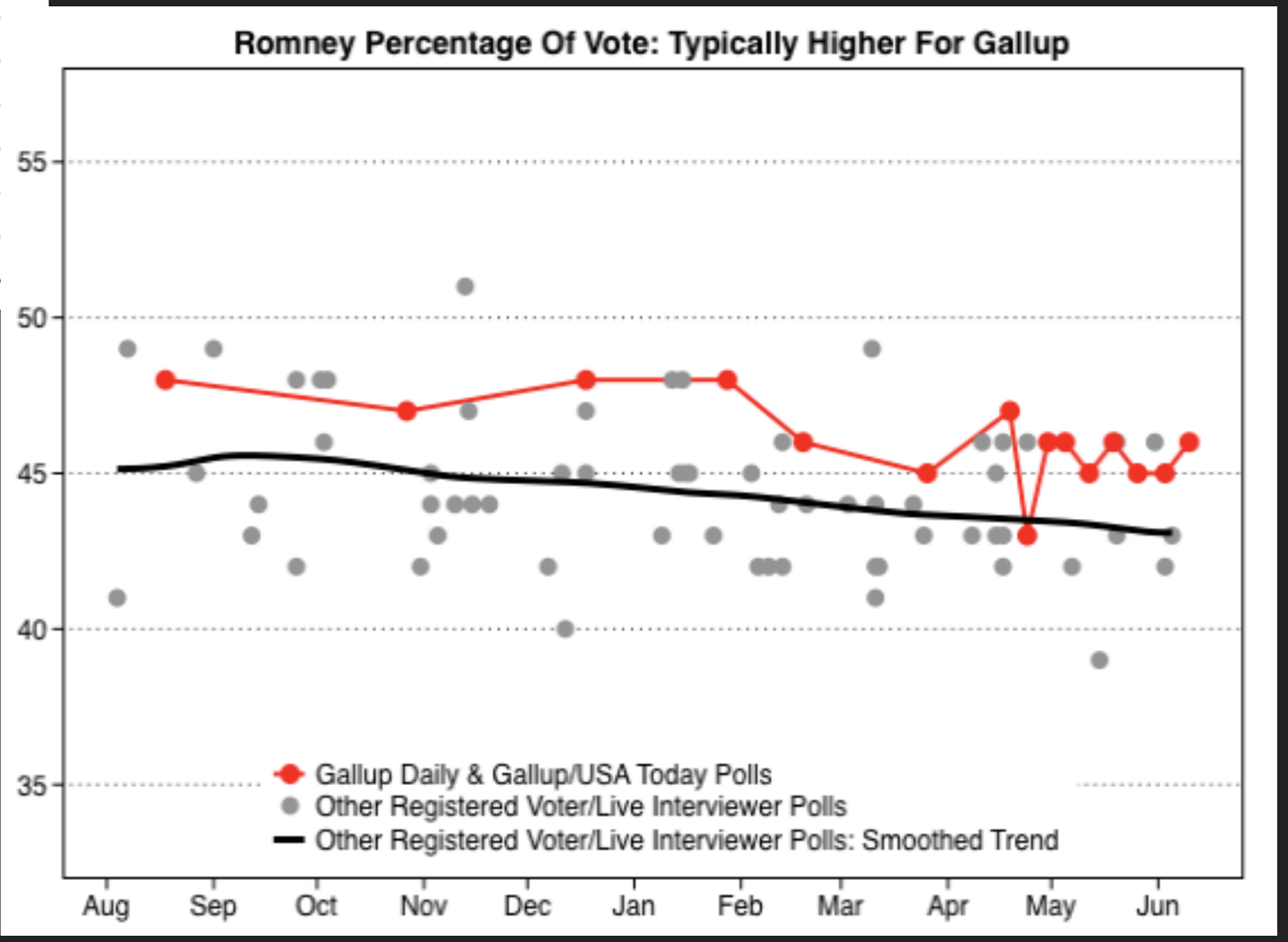
How four pollsters, and The Upshot, interpreted 867 poll responses:

SOURCES OF ERROR: DEMOGRAPHIC WEIGHTS

Pollster	Result	White	Hisp.	Black	Sample
Charles Franklin Marquette Law	Clinton +3	68%	15%	10%	+5 Dem.
Patrick Ruffini Echelon Insights	Clinton +1	67%	14%	12%	+1 Dem.
Omero, Green, Rosenblatt Penn Schoen Berland Research	Clinton +4	65%	15%	12%	+4 Dem.
Corbett-Davies, Gelman, Rothschild Stanford University/Columbia University/Microsoft Research	Trump +1	70%	13%	14%	+1 Rep.
NYT Upshot/Siena College	Clinton +1	69%	14%	12%	+1 Rep.

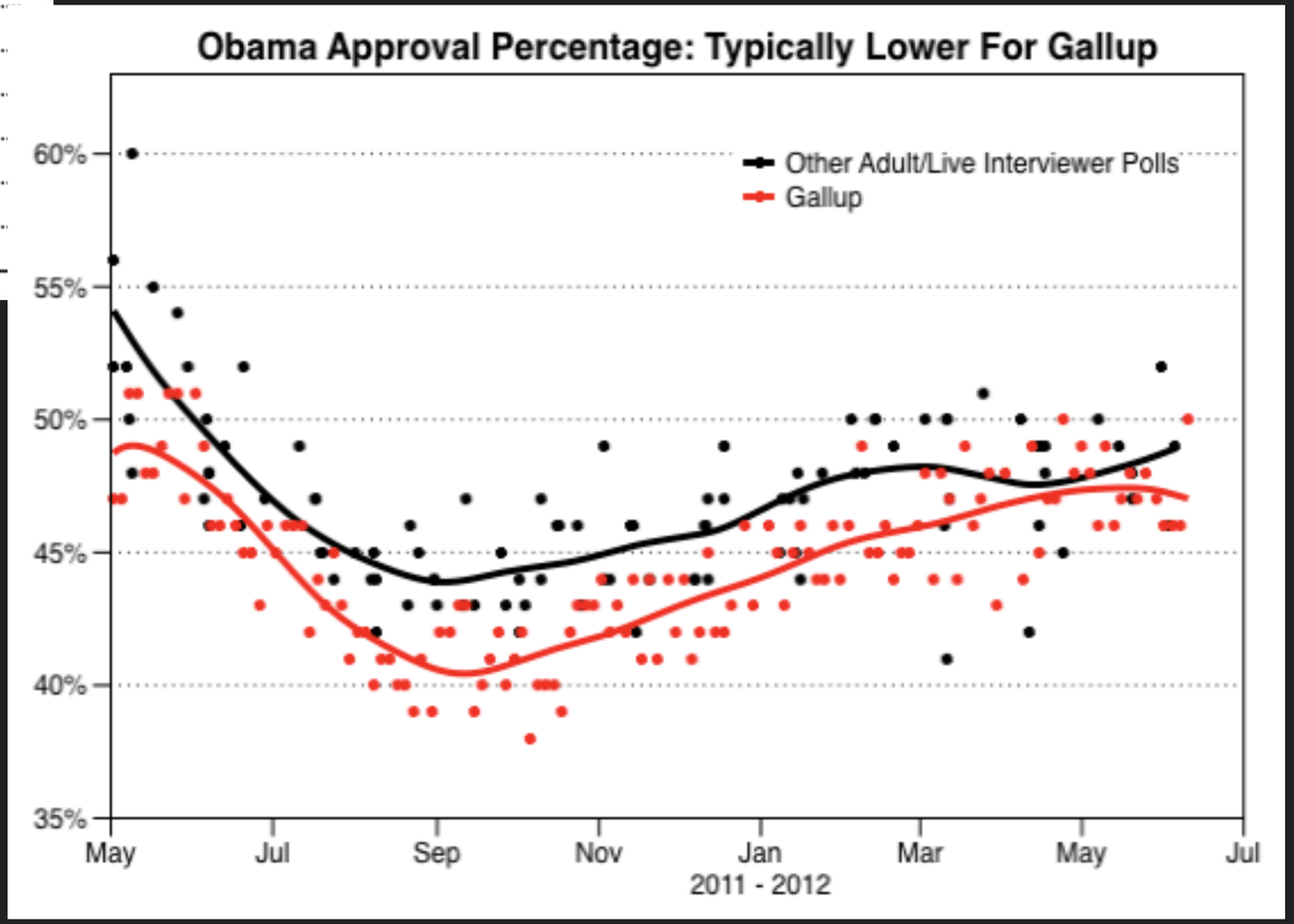
HOUSE EFFECTS

Pollster	House Effect
Pew Research	D +3.2
Public Policy Polling (PPP)	D +3.1
Ipsos	D +2.9
SurveyUSA	D +2.4
Marist (NBC/Marist)	D +1.9
YouGov	D +0.8
CNN (Opinion Research)	D +0.4
Rasmussen Reports	R +1.3
Washington Post / ABC News	R +1.4
Fox News (Robbins & Shaw)	R +1.5
Quinnipiac	R +1.7
Gallup	R +2.5



HOUSE EFFECTS

Pollster	House Effect
Pew Research	D +3.2
Public Policy Polling (PPP)	D +3.1
Ipsos	D +2.9
SurveyUSA	D +2.4
Marist (NBC/Marist)	D +1.9
YouGov	D +0.8
CNN (Opinion Research)	D +0.4
Rasmussen Reports	R +1.3
Washington Post / ABC News	R +1.4
Fox News (Robbins & Shaw)	R +1.5
Quinnipiac	R +1.7
Gallup	R +2.5



POLLING ERROR = UNCERTAINTY

Polls a week before the election aren't perfect predictors

	DEMOCRATIC ADVANTAGE		
YEAR	NATIONAL POLLING AVERAGE	ELECTION RESULT	ABSOLUTE ERROR
1968	-1.2	-0.7	0.5
1972	-25.0	-23.2	1.9
1976	+1.3	+2.1	0.8
1980	-2.5	-9.7	7.2
1984	-17.2	-18.2	1.0
1988	-9.1	-7.7	1.4
1992	+5.7	+5.6	0.1
1996	+11.8	+8.5	3.3
2000	-2.9	+0.5	3.4
2004	-1.6	-2.5	0.9
2008	+7.6	+7.3	0.3
2012	+1.2	+3.9	2.7
Average			**2.0**

POLLING ERROR = UNCERTAINTY

Simple Average Error for polls in the 21 days before an election

	PRESIDENTIAL		STATE-LEVEL		
YEAR	PRIMARY	GENERAL	GOVERNOR	U.S. SENATE	U.S. HOUSE
1998	–	–	8.2	6.8	6.8
2000	7.8	4.5	3.7	5.5	4.5
2002	–	–	5.4	4.5	5.6
2004	7.2	3.2	4.3	5.1	5.0
2006	–	–	4.6	4.3	5.8
2008	7.4	3.4	4.6	5.0	5.8
2010	–	–	4.9	5.4	6.5
2012	8.7	3.6	4.6	4.9	4.8
2014	–	–	4.5	5.4	7.9
2016	9.4	–	–	–	–
All years	**8.1**	**3.6**	**5.1**	**5.1**	**6.4**

FIVETHIRTYEIGHT

POLLING ERROR = UNCERTAINTY

Share of polls that correctly picked the winner

	PRESIDENTIAL		STATE-LEVEL		
YEAR	PRIMARY	GENERAL	GOVERNOR	U.S. SENATE	U.S. HOUSE
1998	–	–	84%	87%	62%
2000	94%	70%	84	84	59
2002	–	–	85	82	82
2004	94	80	80	85	71
2006	–	–	91	92	74
2008	80	92	95	96	84
2010	–	–	85	79	80
2012	61	78	91	87	75
2014	–	–	76	75	91
2016	85	–	–	–	–
All years	**81**	**81**	**85**	**85**	**81**

FIVETHIRTYEIGHT

POLLING ERROR = UNCERTAINTY

Statistical bias in polls

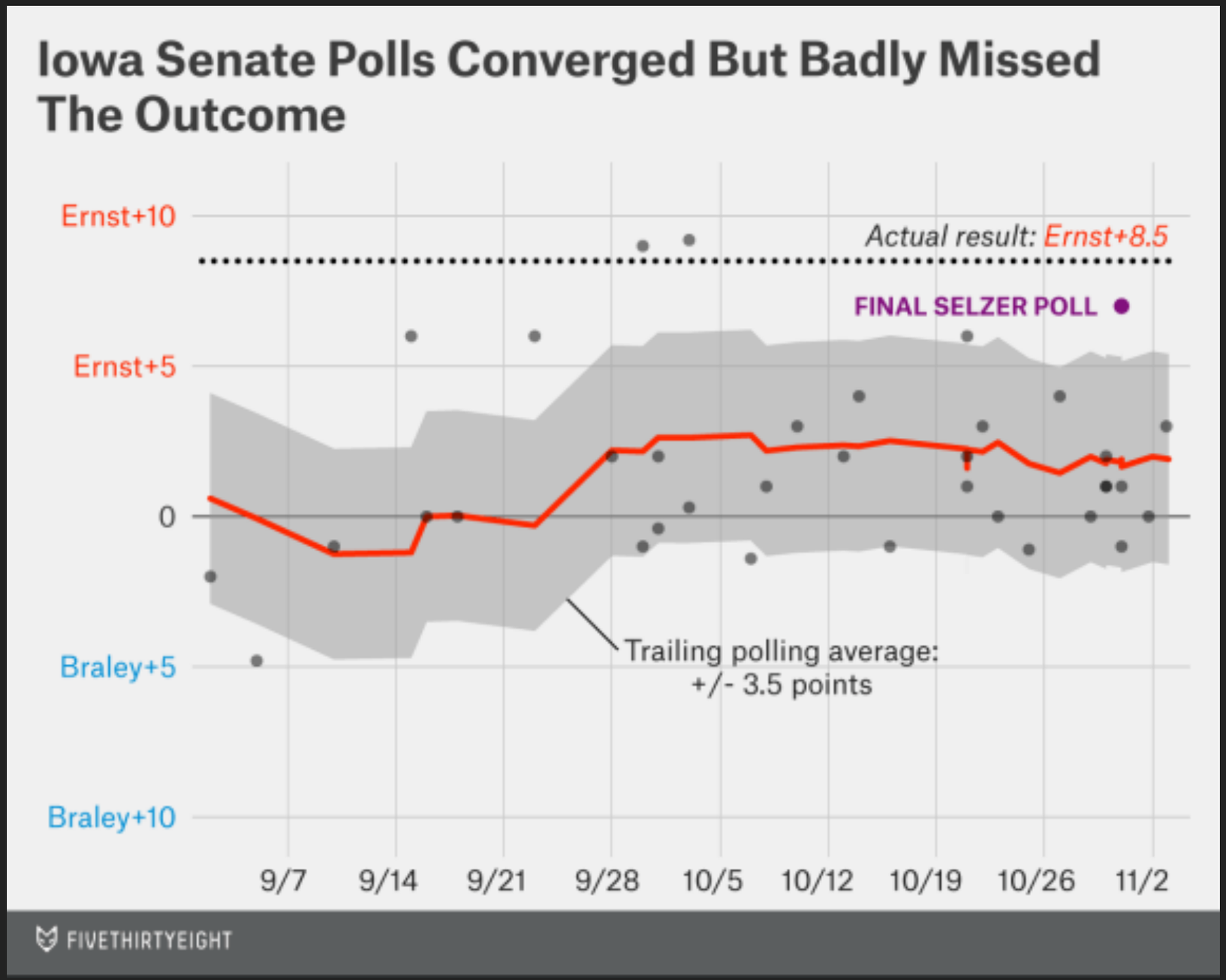
YEAR	PRESIDENTIAL GENERAL	GOVERNOR	U.S. SENATE	U.S. HOUSE
1998		R+6.1	R+4.3	R+1.9
2000	R+2.2	R+0.5	R+2.6	D+1.2
2002		D+3.4	D+1.5	D+2.0
2004	D+1.0	R+1.2	D+0.5	D+2.2
2006		R+0.6	R+2.0	D+0.3
2008	D+0.1	R+1.5	D+0.7	D+1.4
2010		R+1.3	R+2.4	D+1.0
2012	R+2.5	R+2.4	R+3.4	R+2.6
2014		D+2.7	D+3.0	D+3.3
All years	R+0.9	R+0.0	R+1.0	D+1.5

FIVETHIRTYEIGHT

MANAGING UNCERTAINTY: AVERAGES

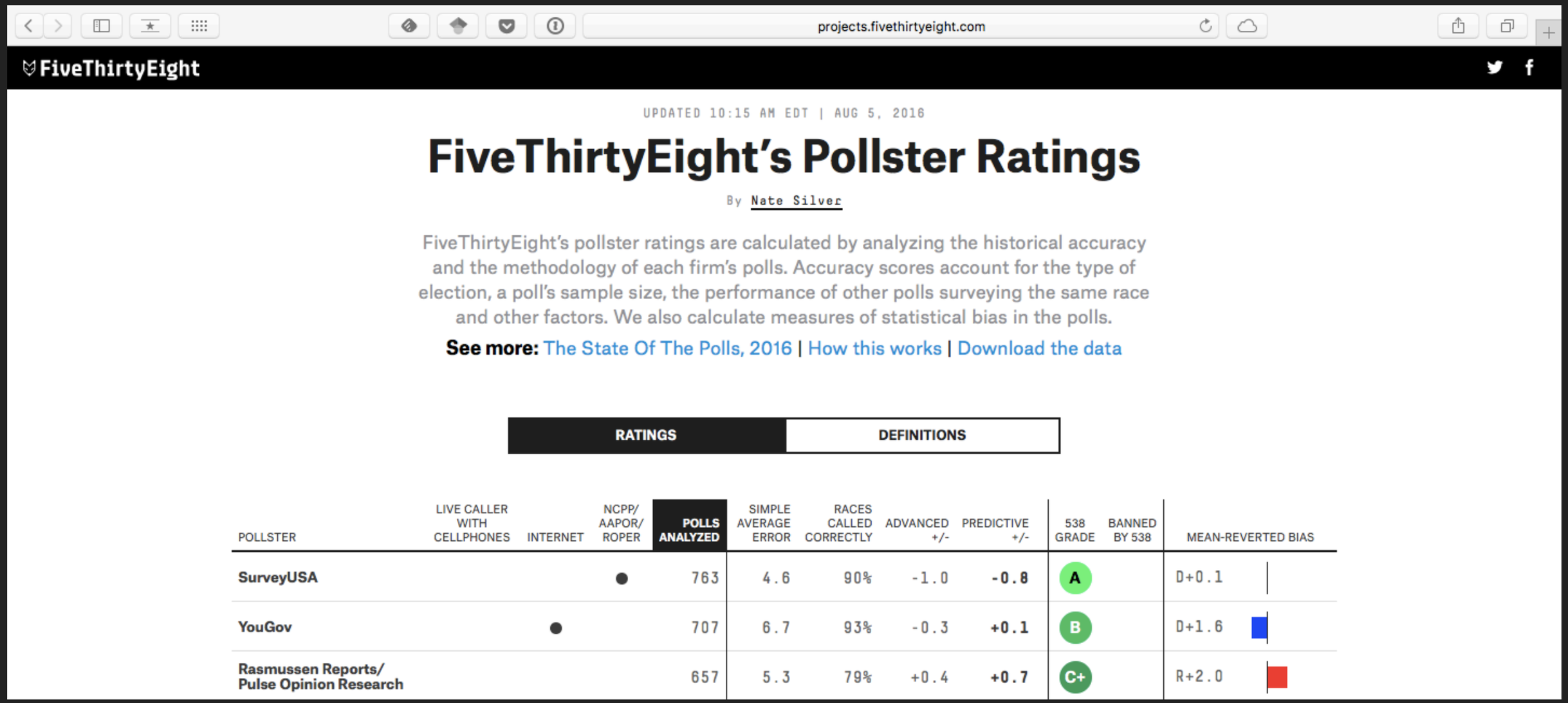
Polling Data						
Poll	**Date**	**Sample**	**MoE**	**Obama (D)**	**Romney (R)**	**Spread**
Final Results	**--**	**--**	**--**	**51.1**	**47.2**	**Obama +3.9**
RCP Average	**10/31 - 11/5**	**--**	**--**	**48.8**	**48.1**	**Obama +0.7**
Politico/GWU/Battleground	11/4 - 11/5	1000 LV	3.1	47	47	Tie
Rasmussen Reports	11/3 - 11/5	1500 LV	3.0	48	49	Romney +1
IBD/TIPP	11/3 - 11/5	712 LV	3.7	50	49	Obama +1
CNN/Opinion Research	11/2 - 11/4	693 LV	3.5	49	49	Tie
Gallup	11/1 - 11/4	2700 LV	2.0	49	50	Romney +1
ABC News/Wash Post	11/1 - 11/4	2345 LV	2.5	50	47	Obama +3
Monmouth/SurveyUSA/Braun	11/1 - 11/4	1417 LV	2.6	48	48	Tie
NBC News/Wall St. Jrnl	11/1 - 11/3	1475 LV	2.6	48	47	Obama +1
Pew Research	10/31 - 11/3	2709 LV	2.2	50	47	Obama +3

MANAGING UNCERTAINTY: HERDING



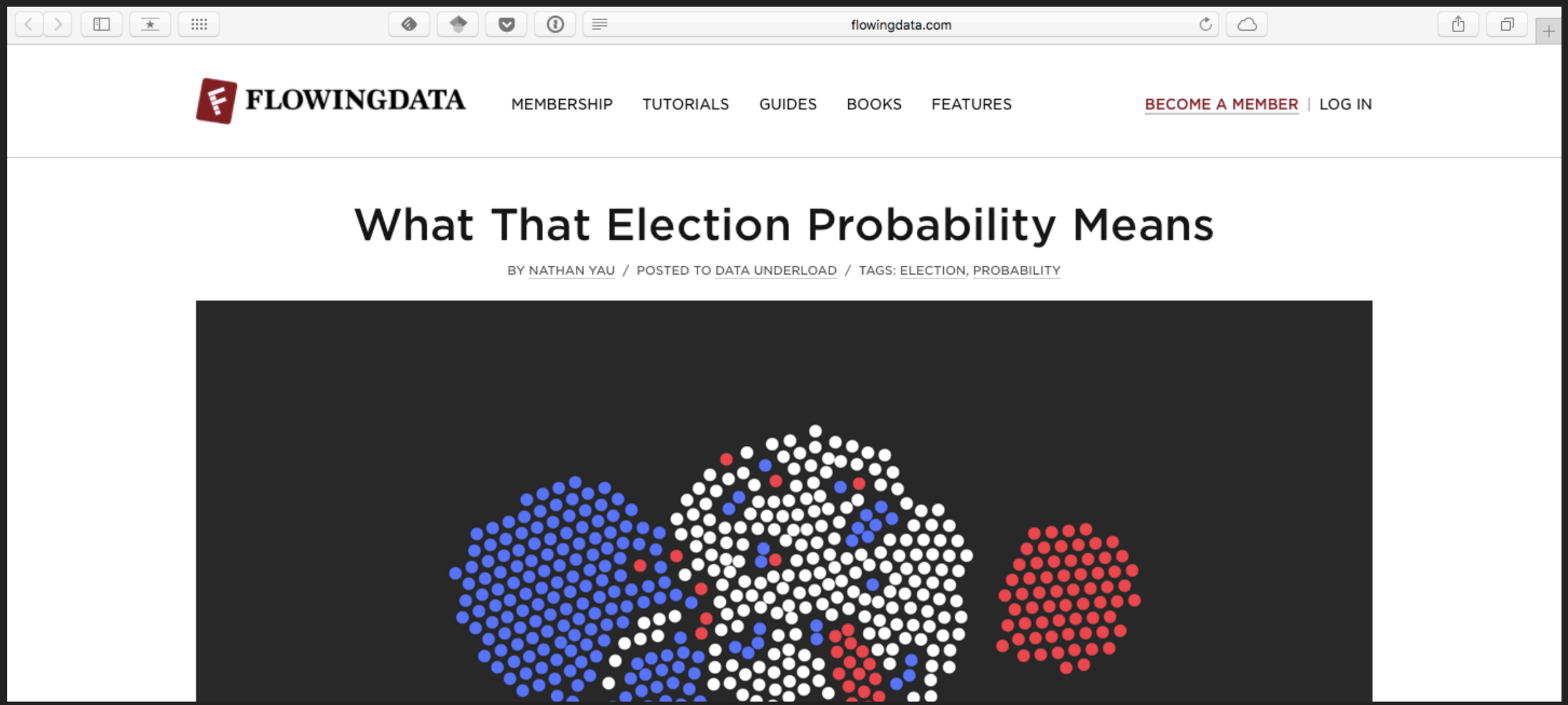
MANAGING UNCERTAINTY: RATING

http://projects.fivethirtyeight.com/pollster-ratings/



MANAGING UNCERTAINTY: RATING

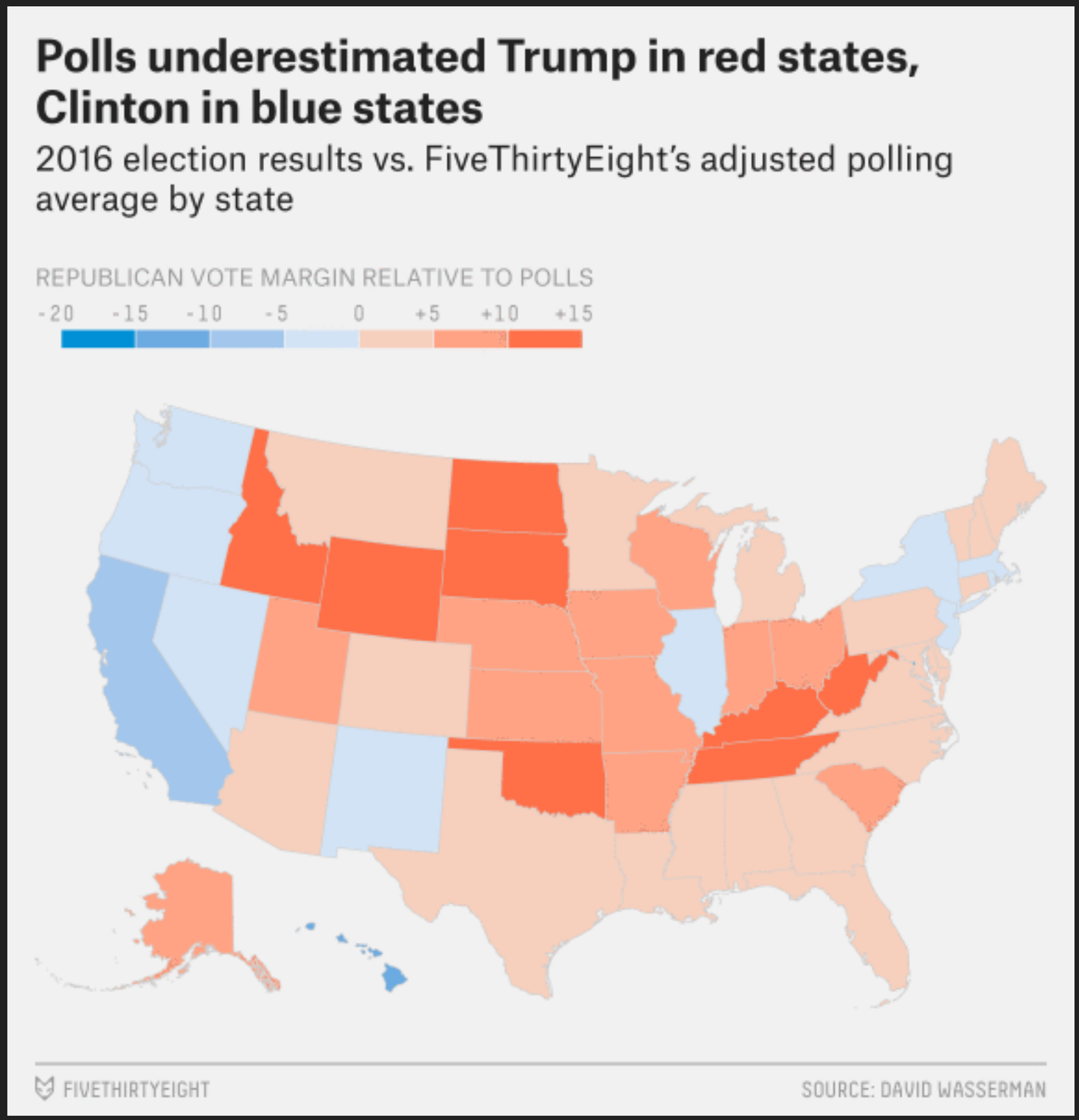
http://flowingdata.com/2016/07/28/what-that-election-probability-means/



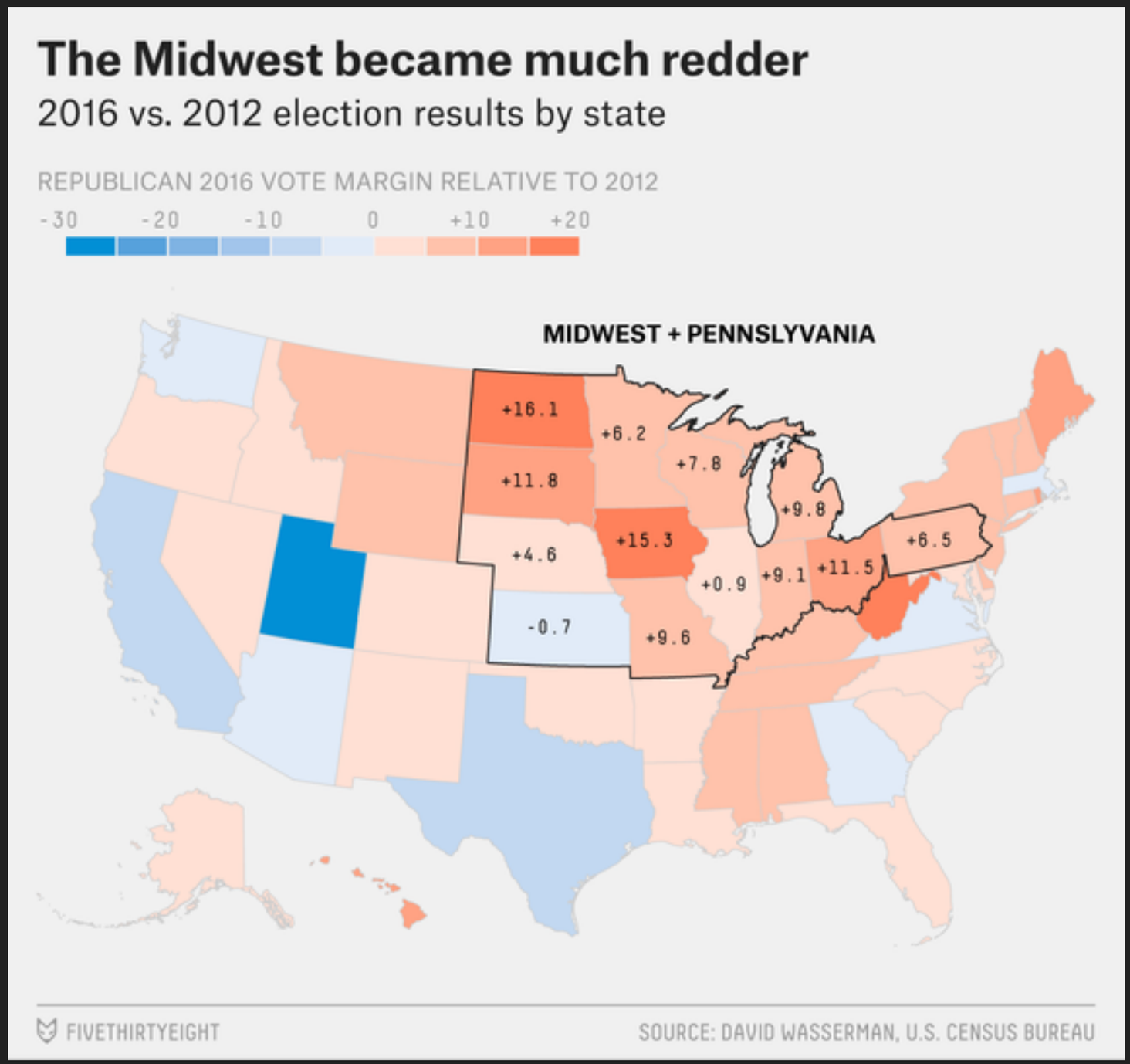
WHAT HAPPENED IN 2016?

Polling Data								
Poll	**Date**	**Sample**	**MoE**	**Clinton (D)**	**Trump (R)**	**Johnson (L)**	**Stein (G)**	**Spread**
Final Results	--	--	--	48.2	46.1	3.3	1.1	Clinton +2.1
RCP Average	11/2 - 11/7	--	--	45.5	42.2	4.7	1.9	Clinton +3.3
Bloomberg	11/4 - 11/6	799 LV	3.5	44	41	4	2	Clinton +3
IBD/TIPP Tracking	11/4 - 11/7	1026 LV	3.1	43	45	8	2	Trump +2
Economist/YouGov	11/4 - 11/7	3677 LV	--	45	41	5	2	Clinton +4
ABC/Wash Post Tracking	11/3 - 11/6	2220 LV	2.5	47	43	4	1	Clinton +4
FOX News	11/3 - 11/6	1295 LV	2.5	48	44	3	2	Clinton +4
Monmouth	11/3 - 11/6	748 LV	3.6	50	44	4	1	Clinton +6
Gravis	11/3 - 11/6	16639 RV	1.0	47	43	3	2	Clinton +4
NBC News/Wall St. Jrnl	11/3 - 11/5	1282 LV	2.7	44	40	6	2	Clinton +4
Reuters/Ipsos	11/2 - 11/6	2196 LV	2.3	42	39	6	3	Clinton +3
Rasmussen Reports	11/2 - 11/6	1500 LV	2.5	45	43	4	2	Clinton +2
CBS News	11/2 - 11/6	1426 LV	3.0	45	41	5	2	Clinton +4

WHAT HAPPENED IN 2016?



WHAT HAPPENED IN 2016?



WHAT HAPPENED IN 2016?

Voters who decided in the final week went strongly for Trump

	VOTE SHARE OF THOSE WHO DECIDED THE WEEK BEFORE THE ELECTION	
STATE	CLINTON	TRUMP
Wisconsin	30%	59%
Minnesota	31	53
Utah	19	41
Iowa	34	54
Pennsylvania	37	54
Florida	38	55
Maine	33	49
New Hampshire	37	52
Michigan	39	50
North Carolina	41	49
New Mexico	41	46
Ohio	43	46
Virginia	45	42
Nevada	45	40
Georgia	52	42

The exit poll did not provide a breakout of voters who decided in the last week in Colorado or Arizona, because the sample size was too small.

SOURCE: NATIONAL EXIT POLL

3 HANDOUTS

PAGE BREAK


```
\newpage
```

INDENTATION

Paragraphs are automatically indented by LaTeX:

```
\par The quick brown fox jumps over the lazy
    sociologist.
```

The *quick* brown fox jumped over the *lazy* sociologist.

REMOVE INDENT

\noindent

Removing the automatic indentation:

```
\par \noindent The quick brown fox jumps over the
    lazy sociologist.
```

The *quick* brown fox jumped over the *lazy* sociologist.

ADD VERTICAL SPACE

`\vspace{val}`

Adding vertical space:

```
\vspace{3mm} \par \noindent The quick brown fox jumps
    over the lazy sociologist.
```

The *quick* brown fox jumped over the *lazy* sociologist.

REMOVING SECTION NUMBERS


```
\section*{Section Title}
```

Removing the section number

```
\section*{Introduction}
The quick brown fox jumped over the lazy sociologist.
```

Introduction

The *quick* brown fox jumped over the *lazy* sociologist.

USING STYLE FILES

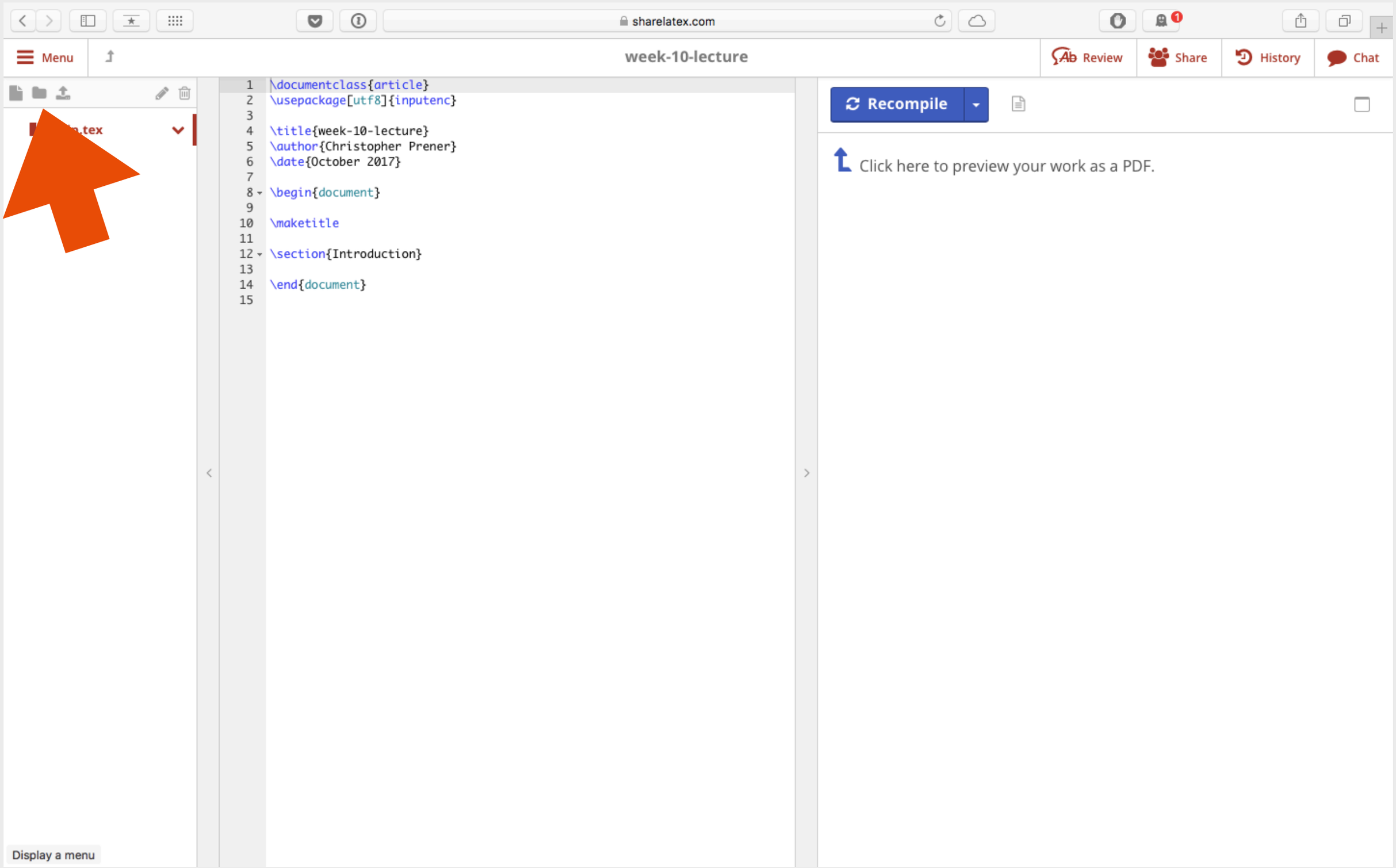
`\usepackage{fileName}`

Using the `genericHandout` style file:

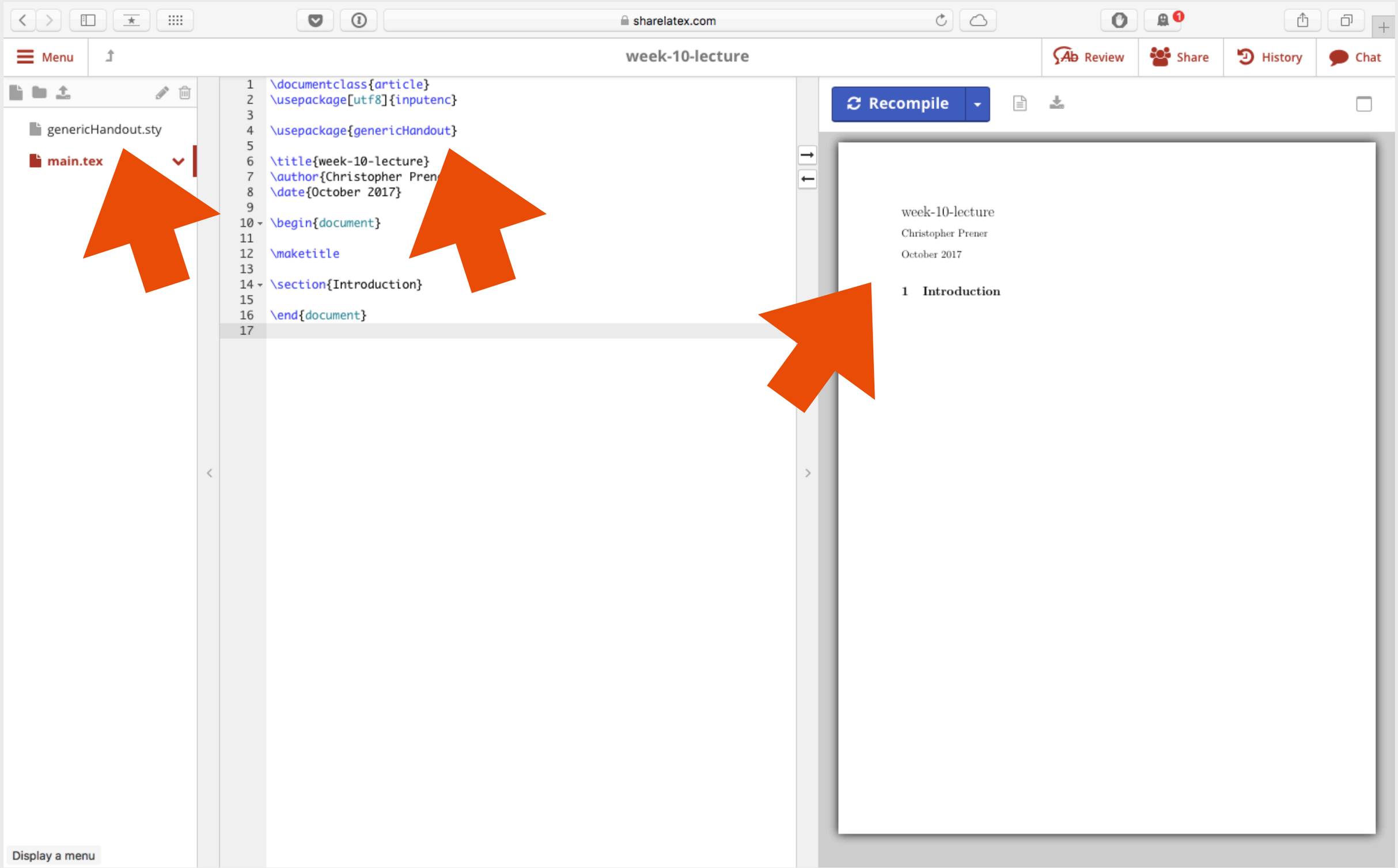
```
\usepackage{genericHandout}
```

The `genericHandout` file *must* be saved in the top-level of your project.

UPLOADING FILES TO SHARELATEX



UPLOADING FILES TO SHARELATEX



LAYOUT

- On page 1:
  - a **title** block with the paper's title and authorship information
  - a **footnote** with conference information
  - an **abstract**, typically of no more than 250 to 300 words
  - your **descriptive statistics table**
  - a few key **references** if audience members want to learn more
- One page 2: **your regression table**

MODULAR AUTHOR AFFILIATIONS

Sample Conference Handout*

Christopher Prener, Ph.D.[1]

[1]Department of Sociology and Anthropology, Saint Louis University

October 30, 2017

Abstract

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

MODULAR AUTHOR AFFILIATIONS

Sample Conference Handout*

Christopher Prener, Ph.D.[1] and Joel Jennings, Ph.D.[1]

[1]Department of Sociology and Anthropology, Saint Louis University

October 30, 2017

Abstract

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

MODULAR AUTHOR AFFILIATIONS

Sample Conference Handout*

Christopher Prener, Ph.D.[1], Joel Jennings, Ph.D.[1], and Taylor Braswell, Ph.D.[2]

[1]Department of Sociology and Anthropology, Saint Louis University
[2]Department of Sociology and Anthropology, Northeastern University

October 30, 2017

Abstract

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

TITLE FOOTNOTE

Sample Conference Handout*

Christopher Prener, Ph.D.[1], Joel Jennings, Ph.D.[1], and Taylor Braswell, Ph.D.[2]

[1]Department of Sociology and Anthropology, Saint Louis University
[2]Department of Sociology and Anthropology, Northeastern University

October 30, 2017

Abstract

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

PREAMBLE SET-UP

```
% add ability to include author affiliations
\usepackage{authblk}

% define conference footnote
\def\conference{Paper presented at the 2017 American Sociological
Association Annual Meeting in Montreal, Quebec. Contact Chris at
chris.prener@slu.edu for a copy of the paper.}

% define title
\title{Sample Conference Handout\thanks{\conference}}
\author[1]{Christopher Prener, Ph.D.}
\affil[1]{Department of Sociology and Anthropology, Saint Louis University}
\date{October 30, 2017}
```

PREAMBLE SET-UP

```
% add ability to include author affiliations
\usepackage{authblk}

% define conference footnote
\def\conference{Paper presented at the 2017 American Sociological
Association Annual Meeting in Montreal, Quebec. Contact Chris at
chris.prener@slu.edu for a copy of the paper.}

% define title
\title{Sample Conference Handout\thanks{\conference}}
\author[1]{Christopher Prener, Ph.D.}
\affil[1]{Department of Sociology and Anthropology, Saint Louis University}
\date{October 30, 2017}
```

PREAMBLE SET-UP

```
% add ability to include author affiliations
\usepackage{authblk}

% define conference footnote
\def\conference{Paper presented at the 2017 American Sociological
Association Annual Meeting in Montreal, Quebec. Contact Chris at
chris.prener@slu.edu for a copy of the paper.}

% define title
\title{Sample Conference Handout\thanks{\conference}}
\author[1]{Christopher Prener, Ph.D.}
\affil[1]{Department of Sociology and Anthropology, Saint Louis University}
\date{October 30, 2017}
```

PREAMBLE SET-UP

```
% add ability to include author affiliations
\usepackage{authblk}

% define conference footnote
\def\conference{Paper presented at the 2017 American Sociological
Association Annual Meeting in Montreal, Quebec. Contact Chris at
chris.prener@slu.edu for a copy of the paper.}

% define title
\title{Sample Conference Handout\thanks{\conference}}
\author[1]{Christopher Prener, Ph.D.}
\affil[1]{Department of Sociology and Anthropology, Saint Louis University}
\date{October 30, 2017}
```

4 COVARIANCE

VARIANCE

Second Moment

Let:

- s^2 = variance
- x = value of observation
- $\bar{x}$ = mean of x
- n = sample size

$$s^2 = \frac{\sum_{i=1}^{n} (x - \overline{x})^2}{n - 1}$$

VARIANCE RE-STATED

$$s^2 = \frac{\sum_{i=1}^{n} (x - \overline{x})(x - \overline{x})}{n - 1}$$

COVARIANCE

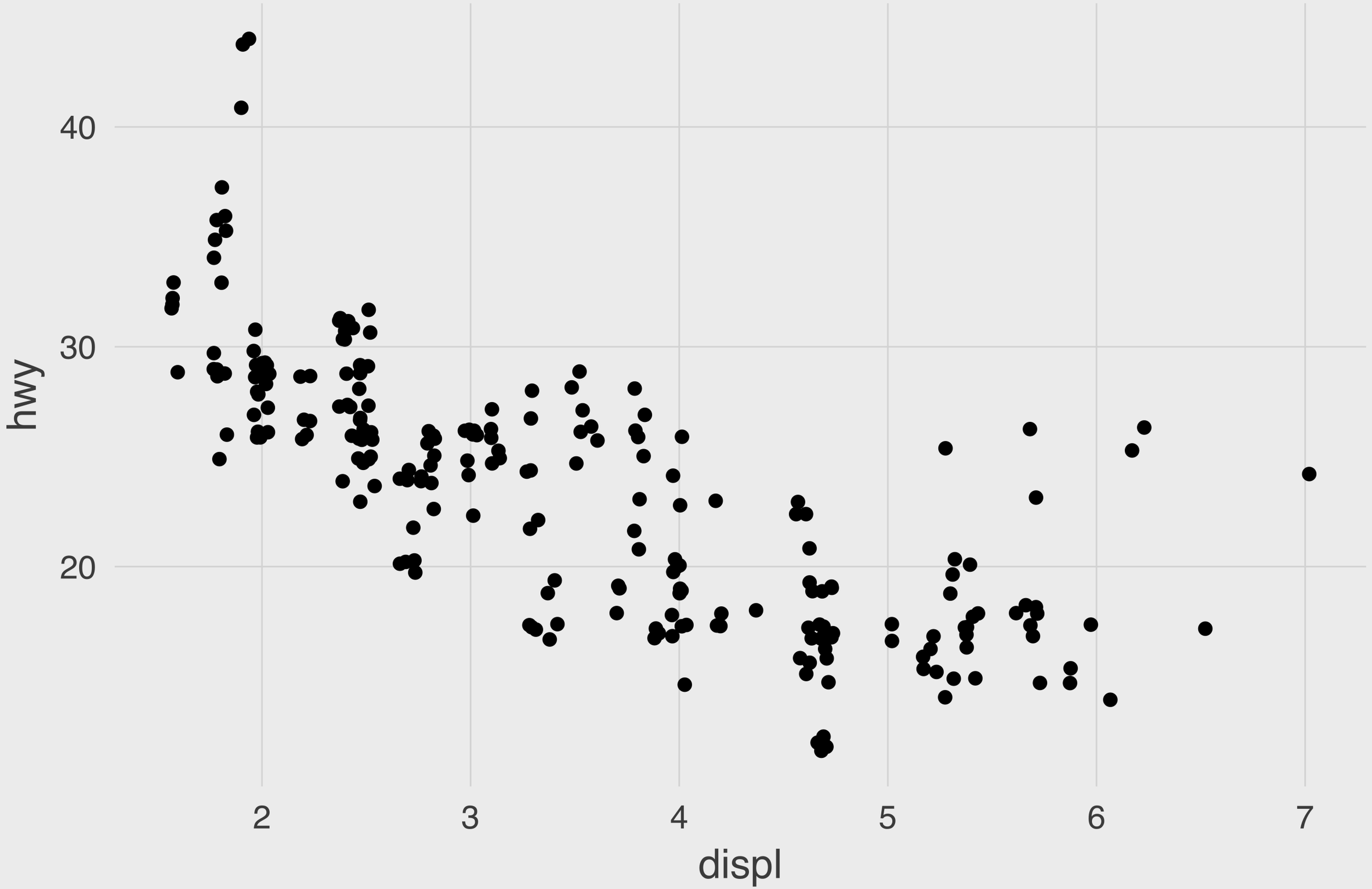
Let:

- s^2 = variance
- x = value of observation
- $\bar{x}$ = mean of x
- y = value of observation
- $\bar{y}$ = mean of y
- n = sample size

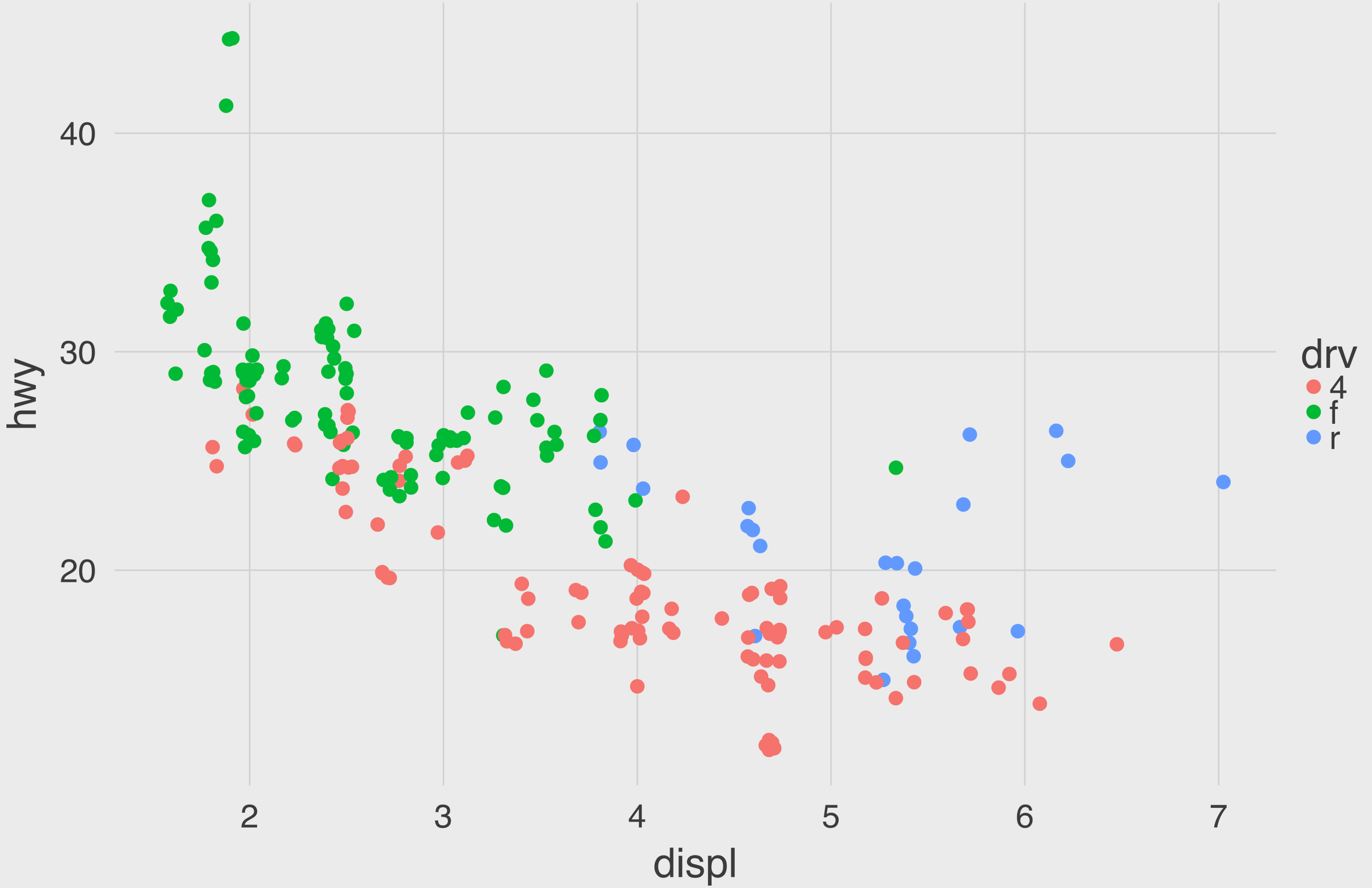
$$s^2 = \frac{\sum_{i=1}^{n} (x - \overline{x})(y - \overline{y})}{n - 1}$$

5 SCATTERPLOTS

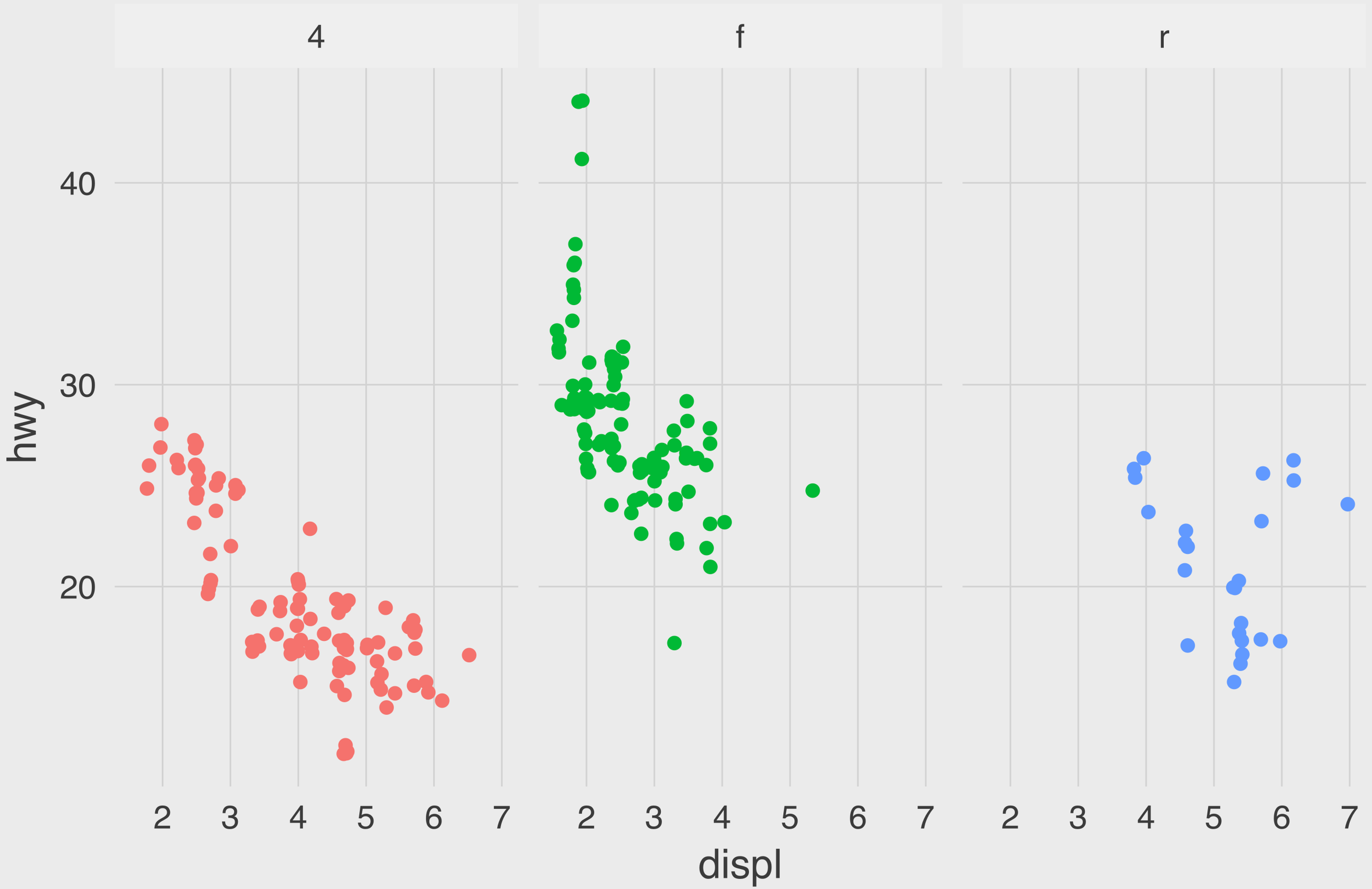
SCATTER PLOT



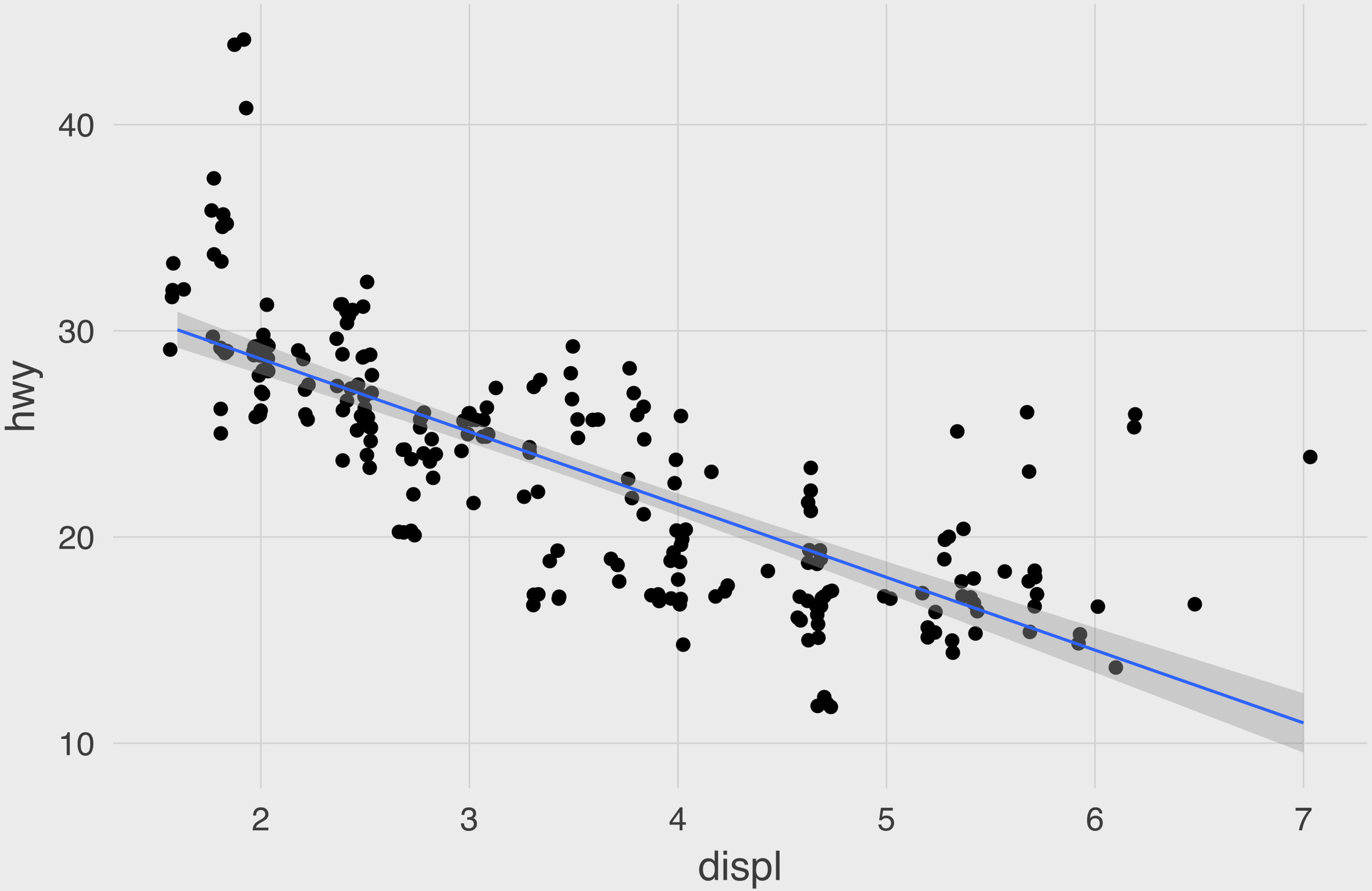
SCATTER PLOT WITH GROUPS



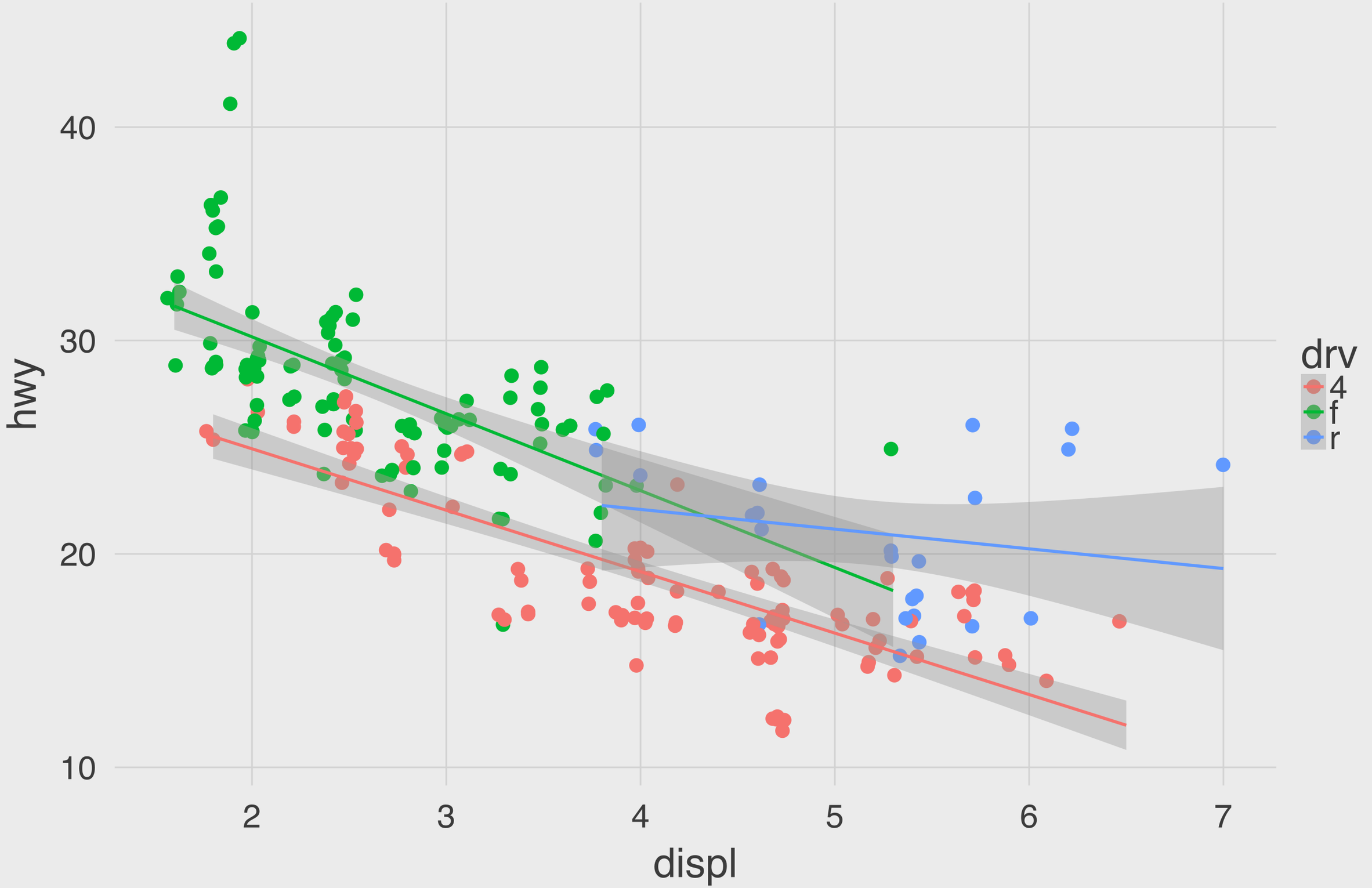
SCATTER PLOT WITH FACETS



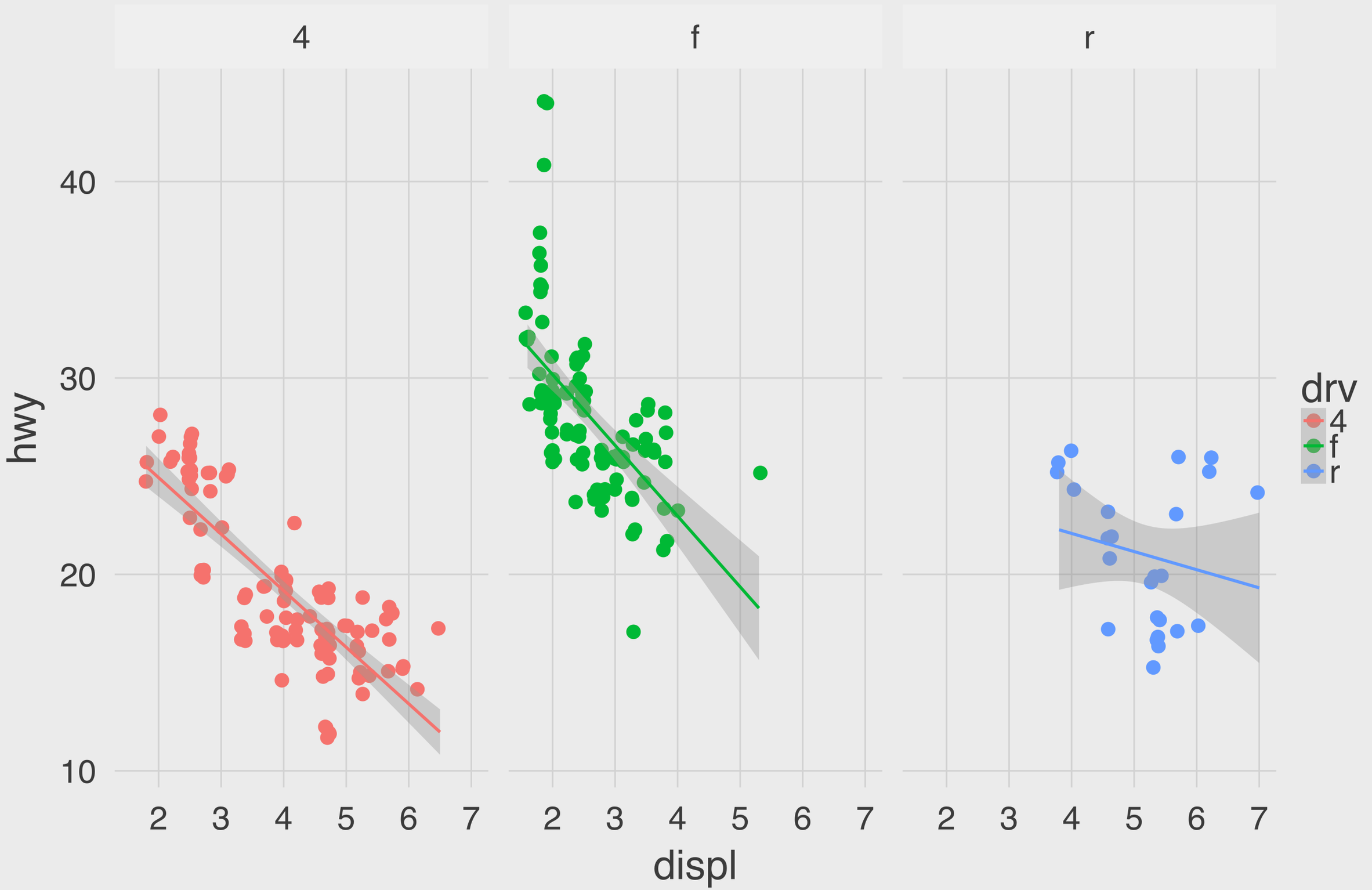
SCATTER PLOT WITH LINE



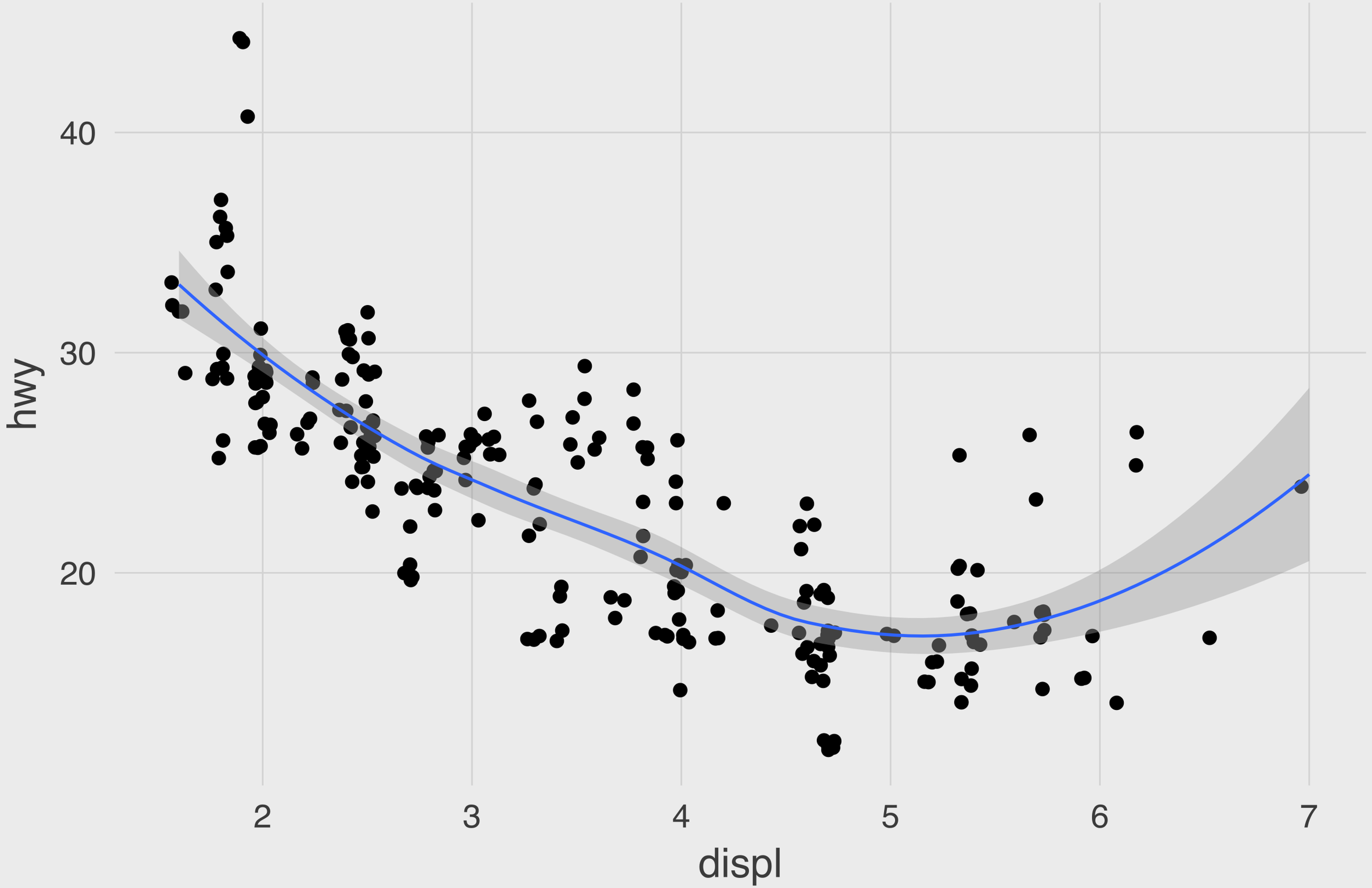
SCATTER PLOT WITH GROUP LINES



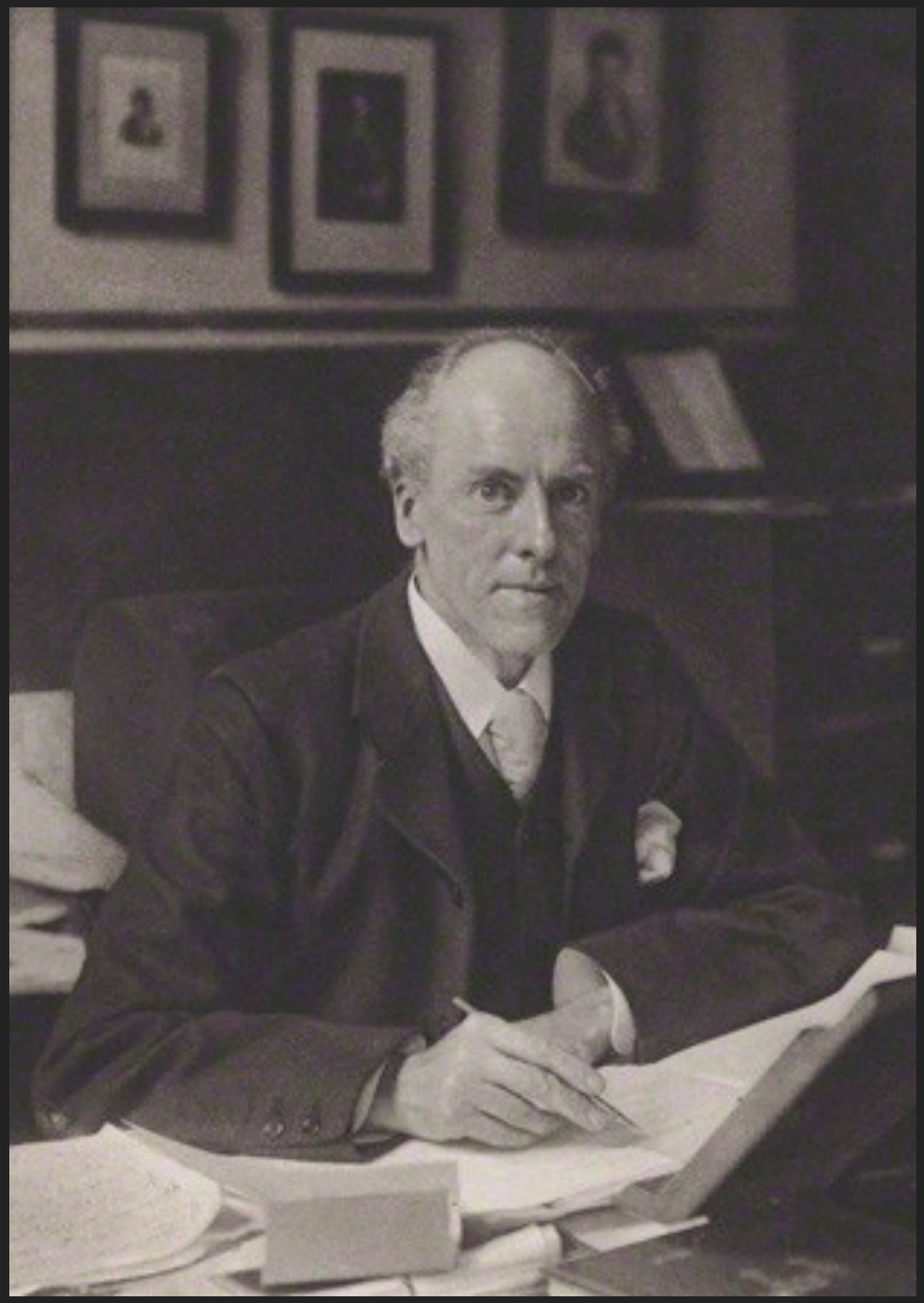
SCATTER PLOT WITH FACETS & GROUP LINES



SCATTER PLOT WITH LOESS



6 PEARSON'S R



KARL PEARSON

- English mathematician
- Student of Sir Francis Galton (and, like Galton, he was a eugenicist and social Darwinist)
- Developed the "product-moment correlation coefficient" based on work Galton had done in 19th century.
- Also introduced "moments", histograms, and core concepts related to statistical significance (including the "p-value")!

COVARIANCE

$$s^2 = \frac{\sum_{i=1}^{n}(x - \overline{x})(y - \overline{y})}{n - 1}$$

CORRELATION

Product of first moments

$$r = \frac{\sum_{i=1}^{n}(x-\overline{x})(y-\overline{y})}{(n-1)s_x s_y}$$

Product-Moment Correlation Coefficient

CORRELATION

Let:

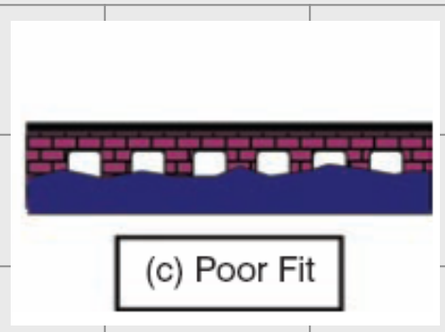
- x = value of observation
- $\bar{x}$ = mean of x
- y = value of observation
- $\bar{y}$ = mean of y
- s_x = std. deviation of x
- s_y = std. deviation of y
- n = sample size

$$r = \frac{\sum_{i=1}^{n} (x - \bar{x})(y - \bar{y})}{(n-1)s_x s_y}$$

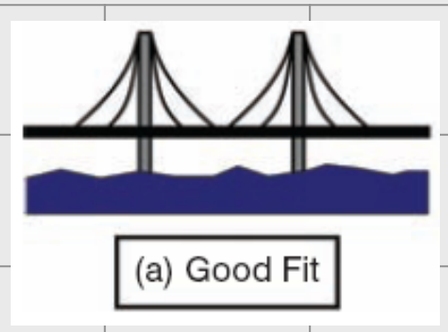
ASSUMPTIONS

- Both x and y should be continuous, normally distributed variables
- There should be a linear relationship between x and y
- Sufficiently large sample size ($n >= 30$)
- There should be no extreme outliers

OUTLIERS & MODEL FIT



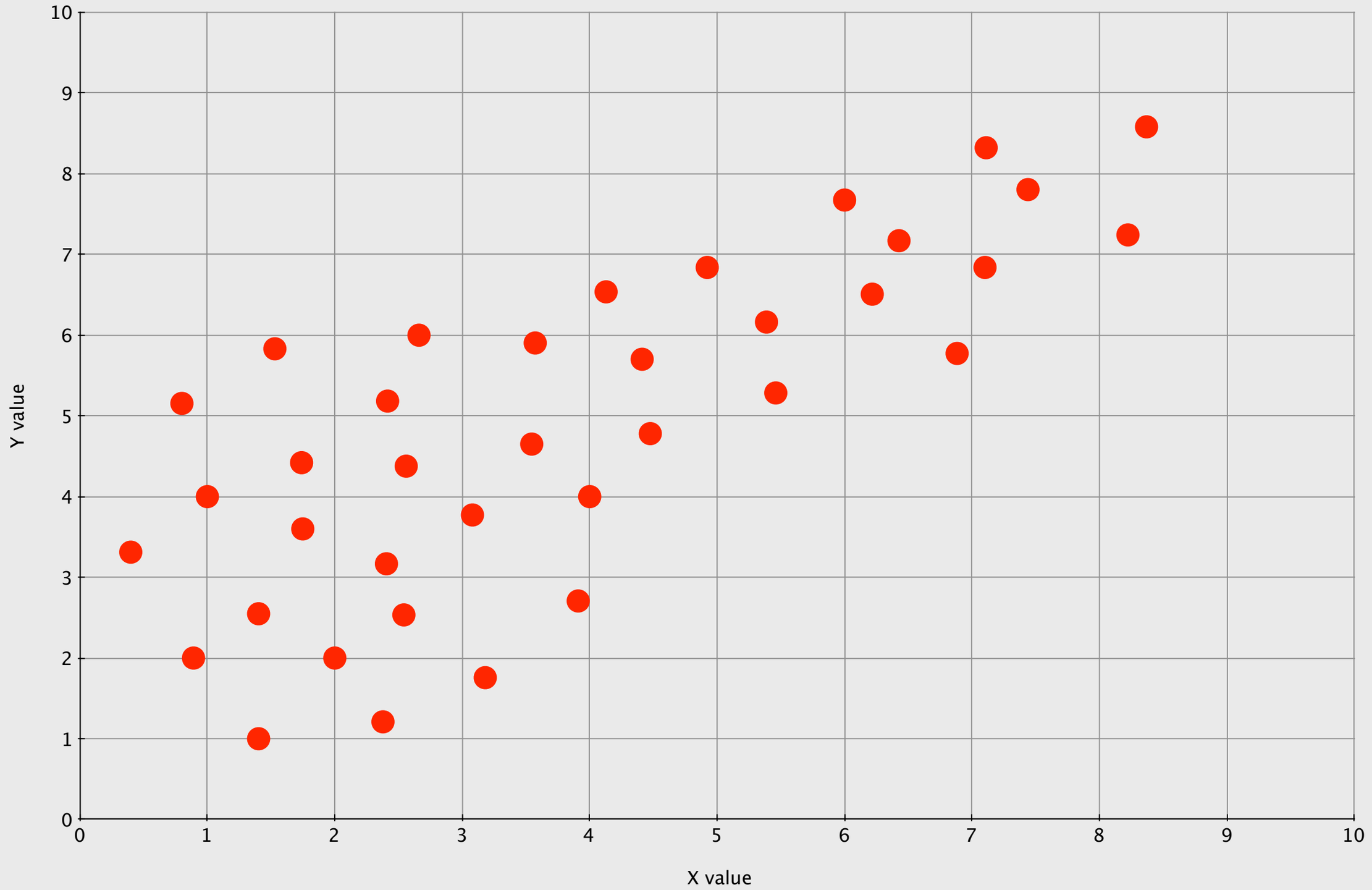
OUTLIERS & MODEL FIT



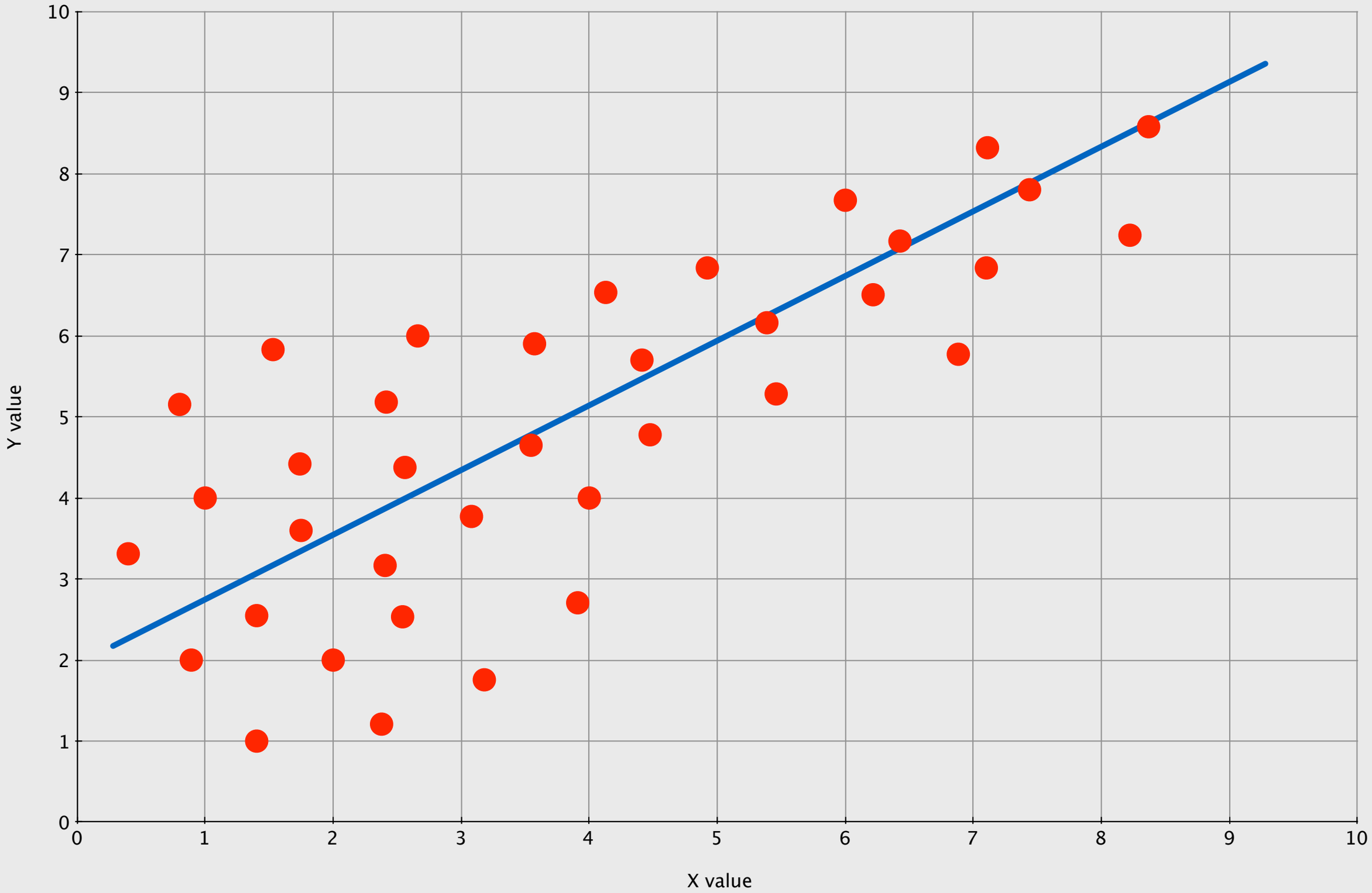
INTERPRETATION

- Range: -1 $<= r <= 1$
- Direction: negative or positive?

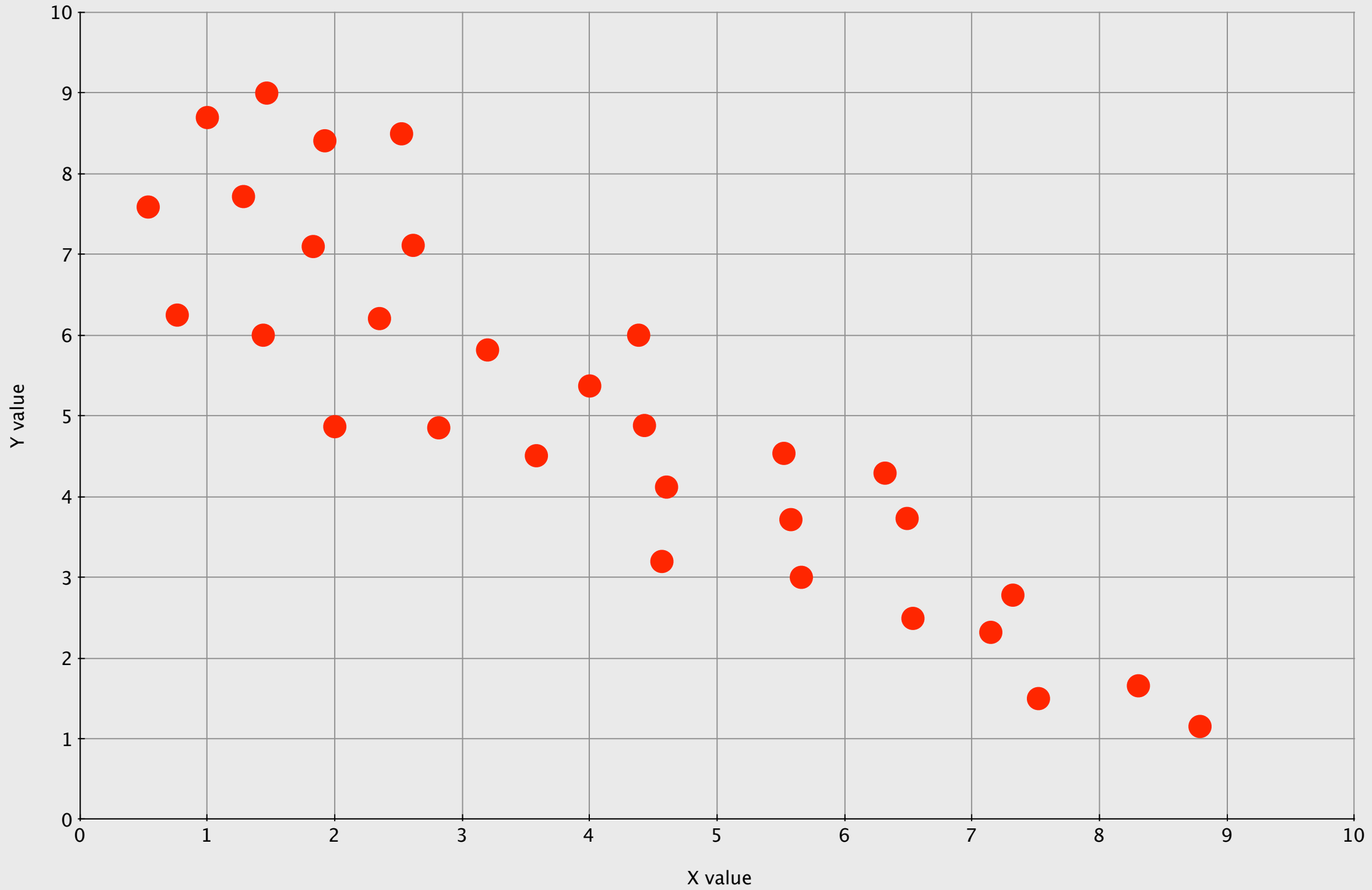
DIRECTION: POSITIVE



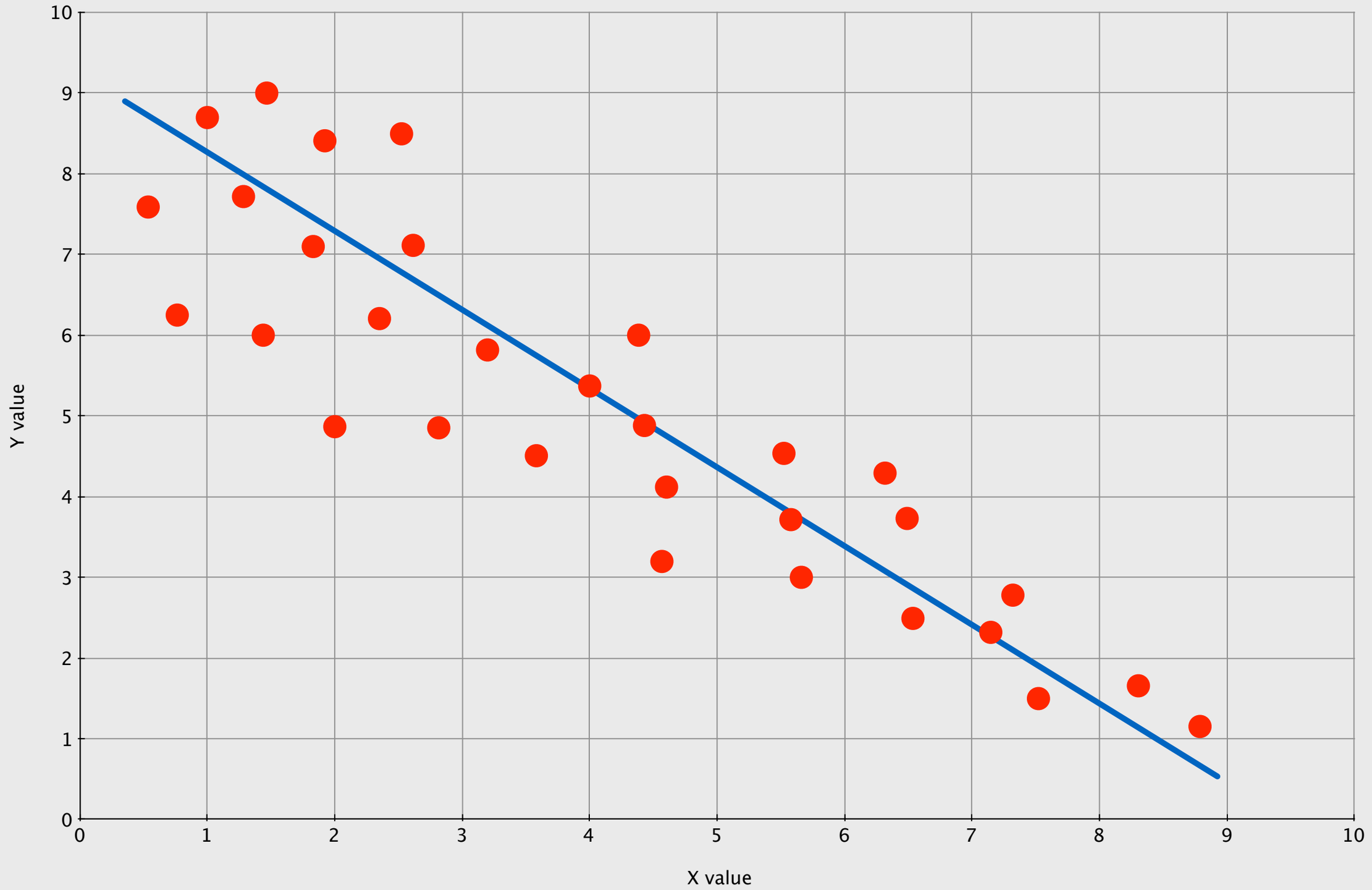
DIRECTION: POSITIVE



DIRECTION: NEGATIVE

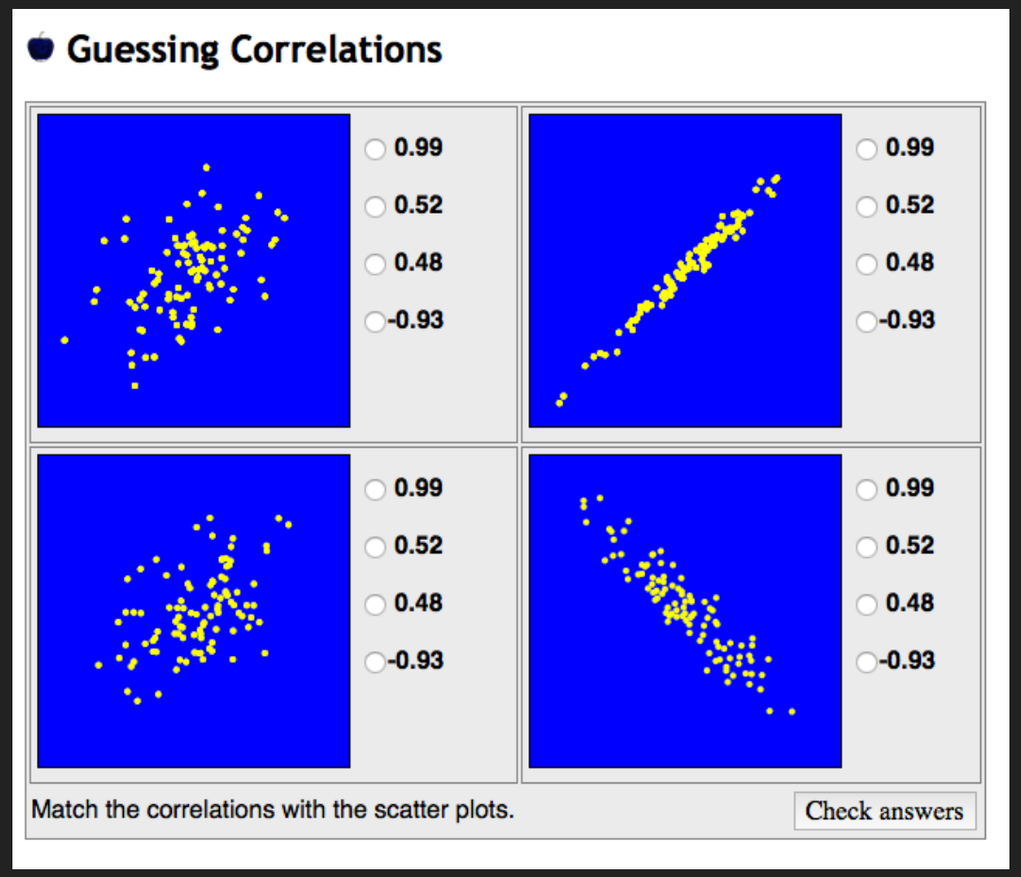


DIRECTION: NEGATIVE



INTERPRETATION

- Range: -1 $<=$ r $<=$ 1
- Direction: negative or positive?
- Effect Size (absolute value):
  - Weak: $0 <= r < .3$
  - Moderate: $.3 <= r < .6$
  - Strong: $.6 <= r < 1$



http://istics.net/Correlations/

PERCENT OF VARIANCE

- Squaring r will give us a value of $0 <= r^2 <= 1$

- r^2 value corresponds to the amount of variation that x accounts for in y (or vice versa)
  - *This is not casual*

- A r^2 value of .8 suggests that x accounts for 80% of the variance in y

$$r^2$$

STATISTICAL SIGNIFICANCE

Let:

- $t = t$ value
- $r =$ correlation coefficient
- $r^2 =$ variance accounted for
- $v =$ degrees of freedom (n-2)

$$t = \frac{r}{\sqrt{\frac{1-r^2}{v}}}$$

Use function from Week-07 to find the p value associated with t.

INTERPRETATION

- Range: -1 $<=$ r $<=$ 1
- Direction: negative or positive?
- Effect Size (absolute value):
  - Weak: $0 <= r < .3$
  - Moderate: $.3 <= r < .6$
  - Strong: $.6 <= r < 1$

x is moderately, negatively correlated (r=-.422; p = .035) with y. x accounts for 17.808% of the variation in y. An increase in x is associated with a small decrease in y [plain English interpretation].

COMMON PITFALLS

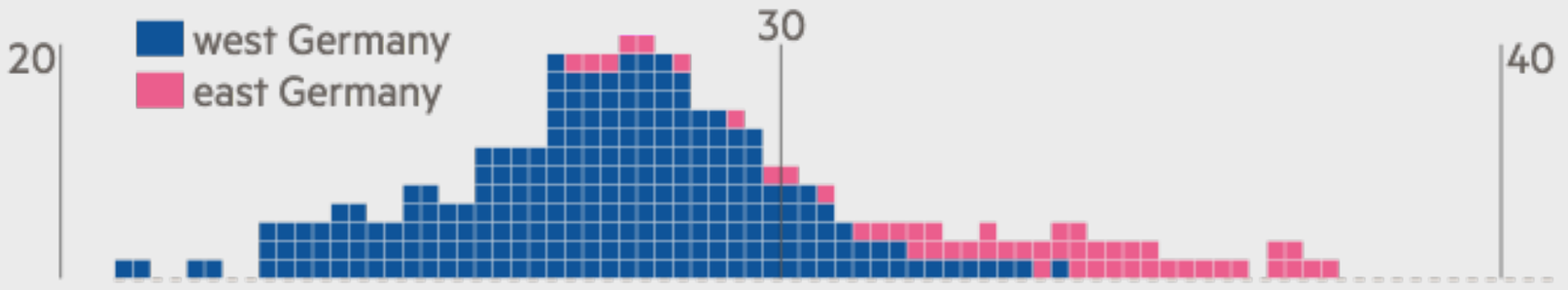
- Correlation != causation, but we should not dismiss strong correlations as not being suggestions of causality either
- $r \approx 0$ does not imply independence (check for non-linearity)
- Ecological fallacy - correlations between groups are stronger than correlations within groups
- Simpson's Paradox - correlations within groups may have different directions than correlation overall

The east/west divide remains imprinted on Germany's demographics, making it key to many of the patterns seen in last week's election

Blocks represent constituencies, stacked to show the spread of demographic characteristics

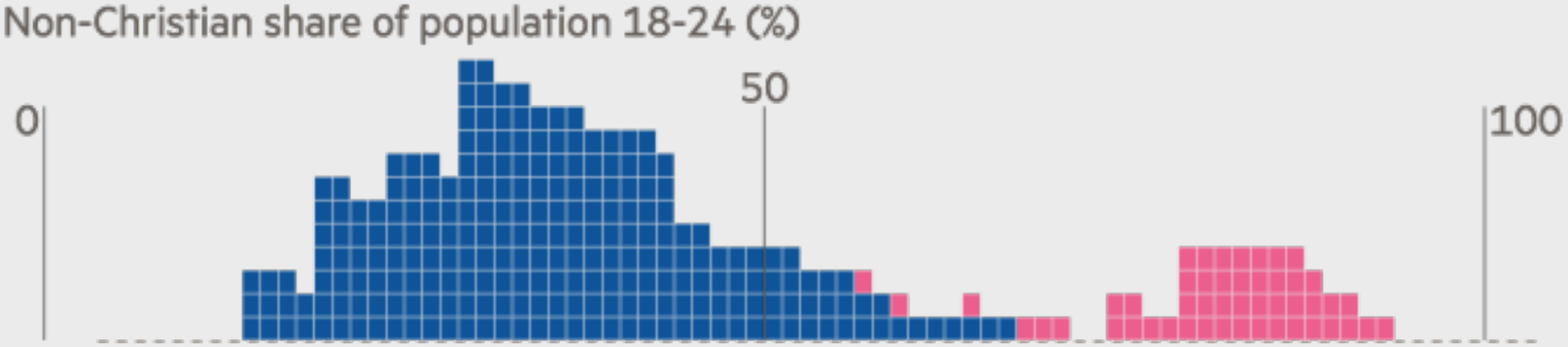
2m young east Germans went west after the wall came down, leaving an old population

Share of population aged 60 and over (%)



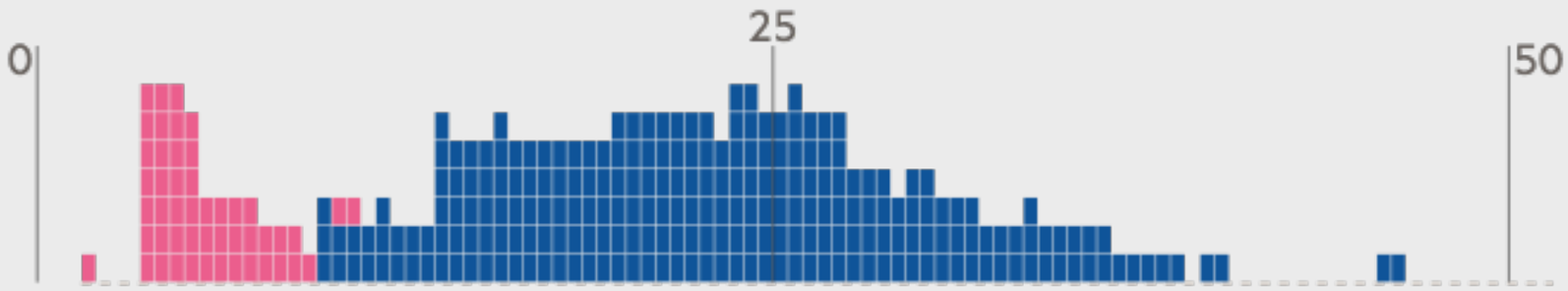
The old East German government restricted Christian churches, leaving a secular region

Non-Christian share of population 18-24 (%)



Far more international migrants have flocked to the west than the east for work

Share of population from a migrant background (%)



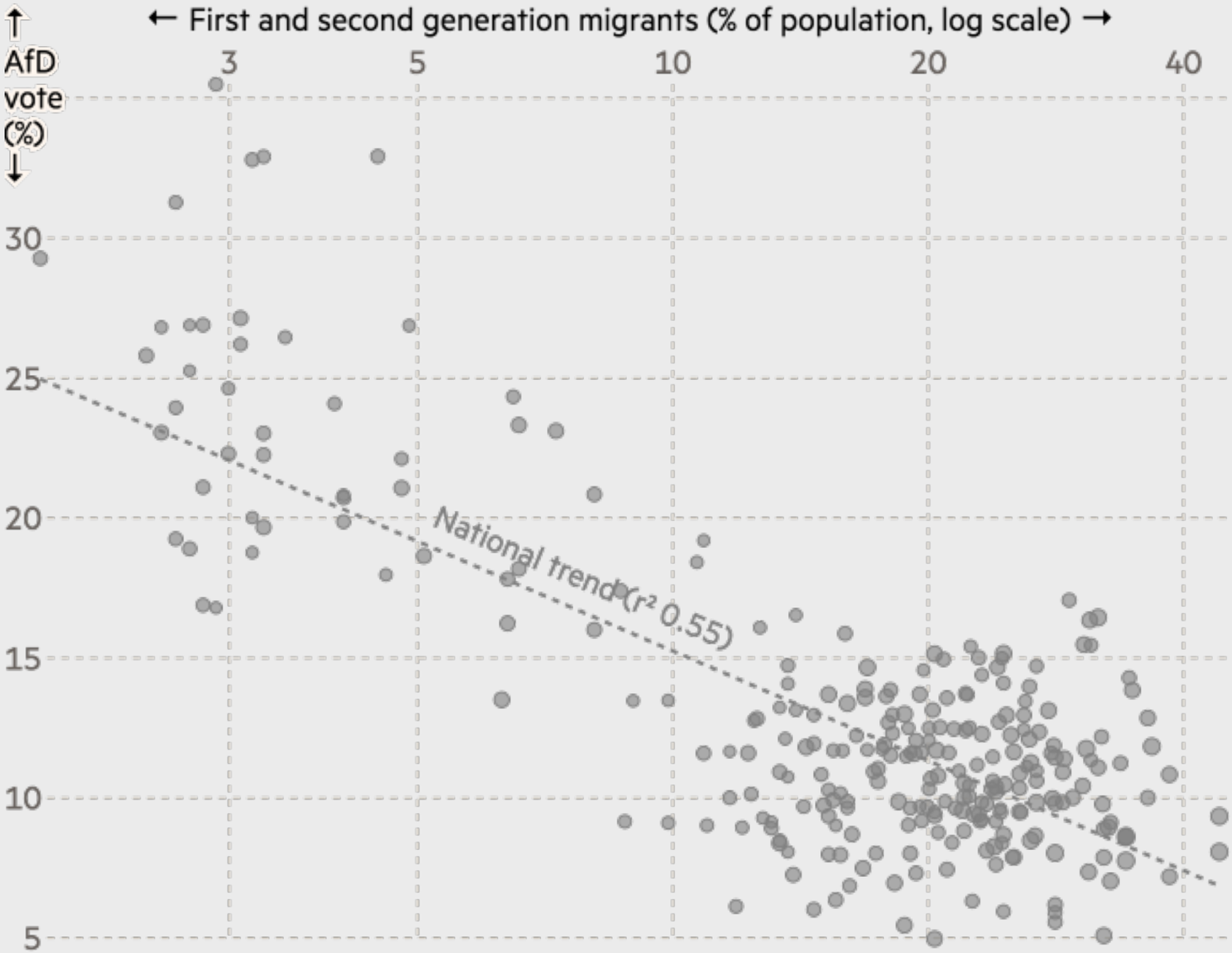
Source: Federal Returning Officer

Graphic: John Burn-Murdoch / @jburnmurdoch

© *FT*

At first glance, AfD fared worse as the number of people from immigrant backgrounds increased

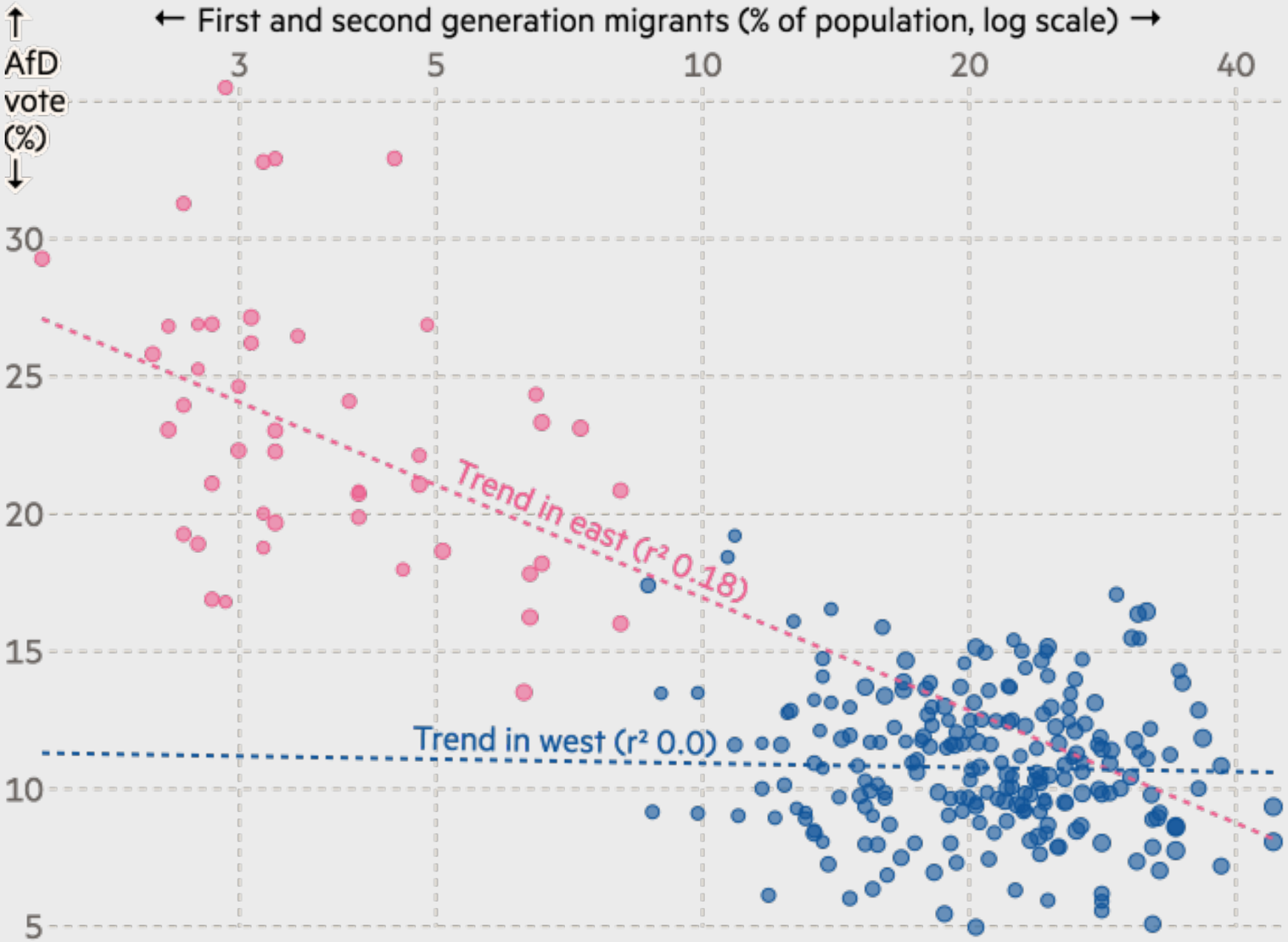
Points represent constituencies



Source: Federal Returning Officer
Graphic: John Burn-Murdoch / @jburnmurdoch
© FT

But when we factor in the east/west divide, the trend in the west disappears completely and we see how reliant the national trend was on the difference between the regions

Points represent constituencies



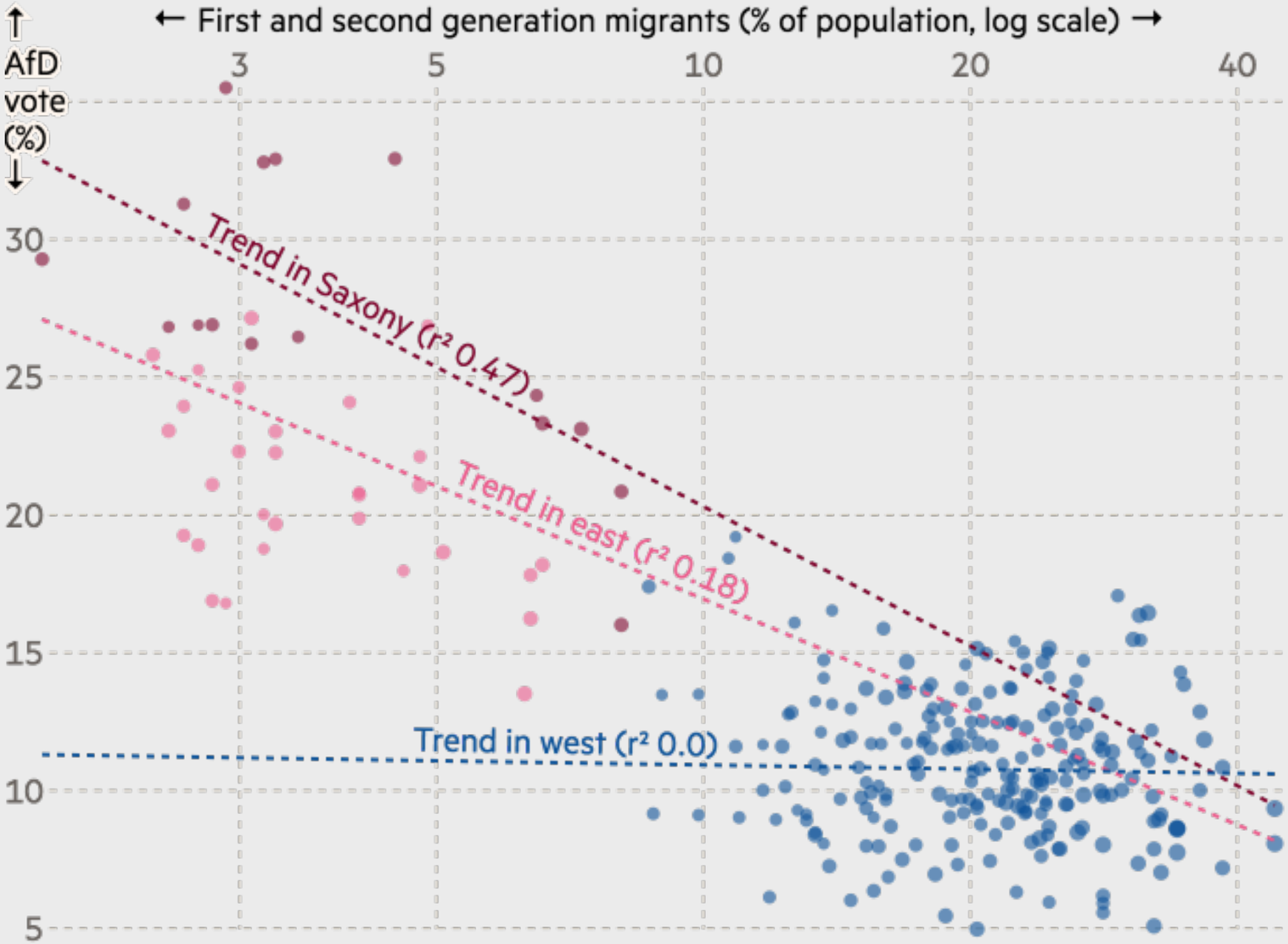
Source: Federal Returning Officer

Graphic: John Burn-Murdoch / @jburnmurdoch

© FT

Adding the context that Saxony has always been a hotbed of right-wing support lays bare another confounding factor that was exaggerating the strength of the national trend

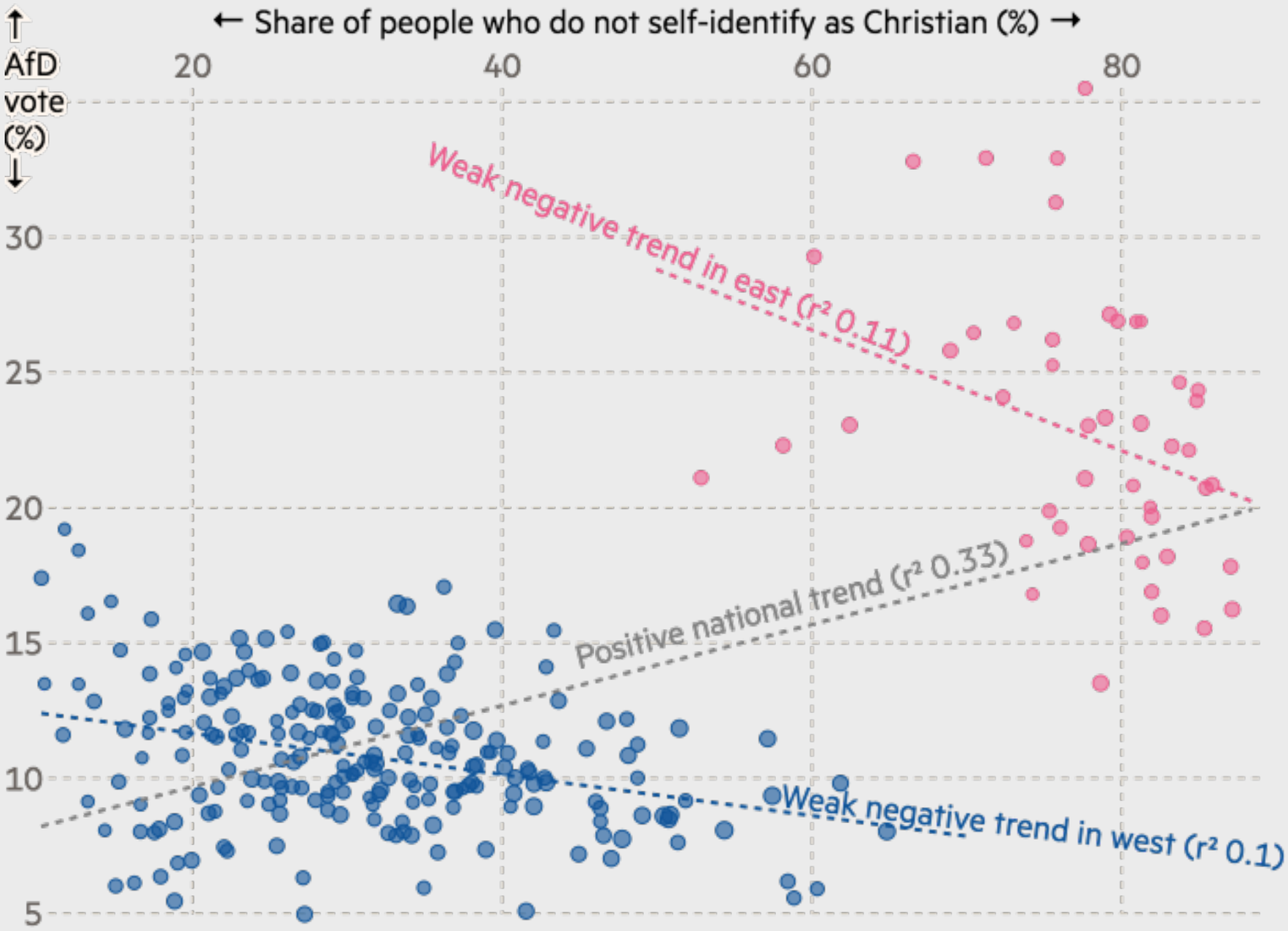
Points represent constituencies



Source: Federal Returning Officer
Graphic: John Burn-Murdoch / @jburnmurdoch
© FT

Danger signs:
The national trend is driven entirely by differences between east and west, and the trend is reversed at regional level

Points represent constituencies



Source: Federal Returning Officer
Graphic: John Burn-Murdoch / @jburnmurdoch

© FT

7 BACK MATTER

WHAT WE COVERED TODAY

2. Public Polling

3. Handouts

4. Covariance

5. Scatterplots

6. Pearson's *r*

REMINDERS

Jotter/wiki has been depreciated - all posts have been migrated to the new website and new content will only be posted there!

Lab-10 due next Monday

Lab-11 and PS-08 due Monday, 11/13

Final project drafts due Monday, 11/13 (everyone should have a handout and draft slides; SOC 5050 students should also have a draft paper to submit)