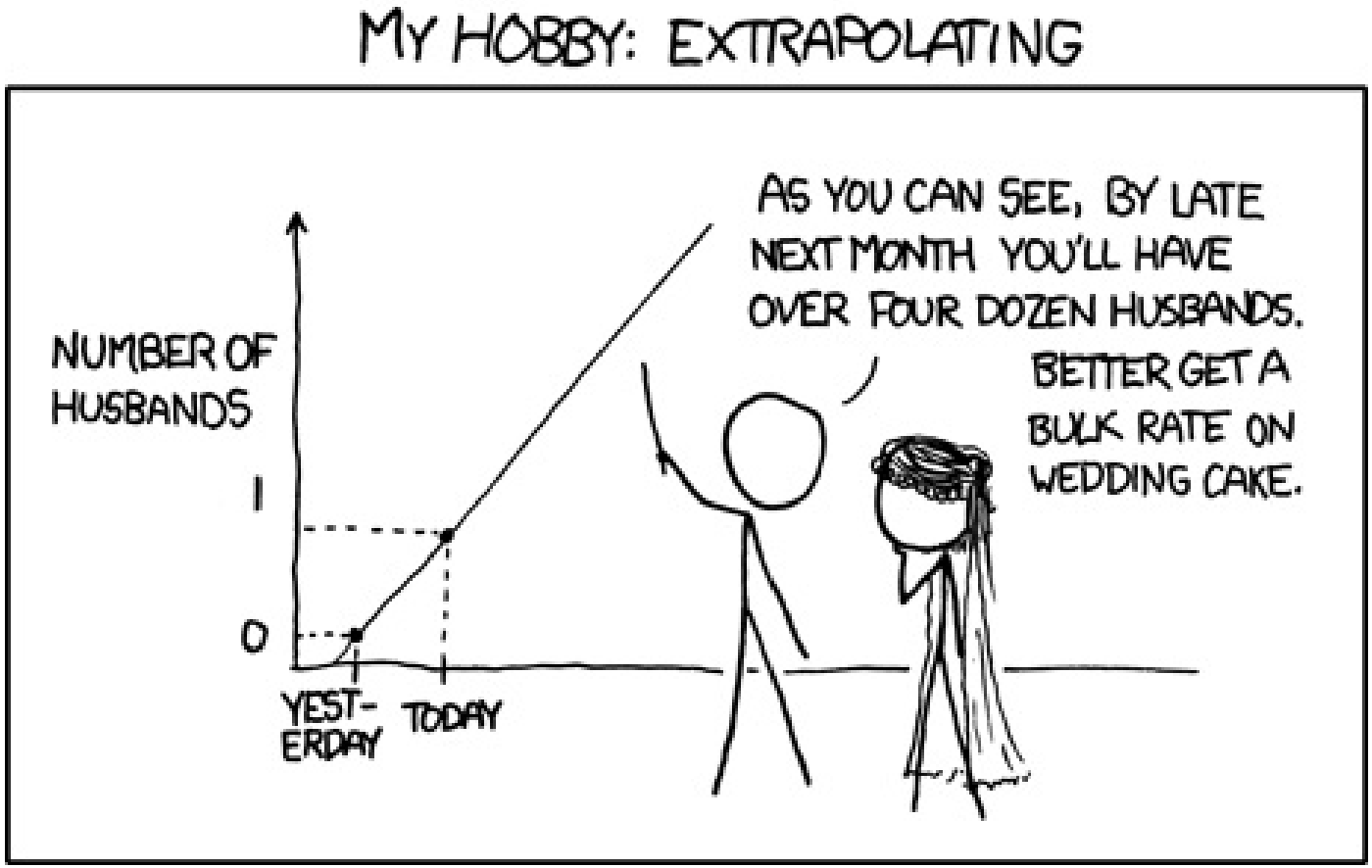
Lecture 14: Regression

Jacob M. Montgomery

Quantitative Political Methodology

An introduction to regression

Correlation and bivariate linear regression

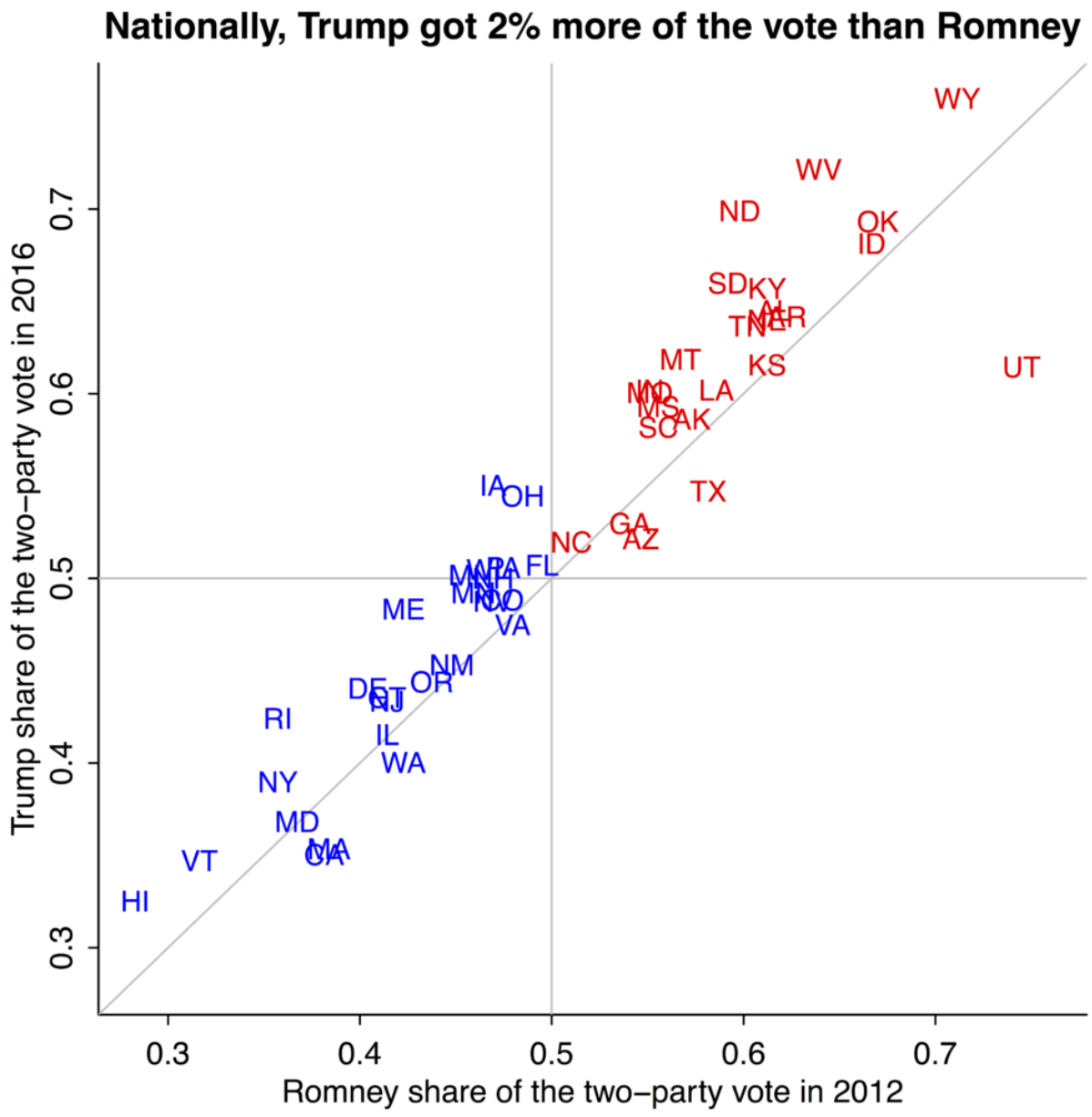


- Scatterplots
- Correlation
- Draing the "best" line through data

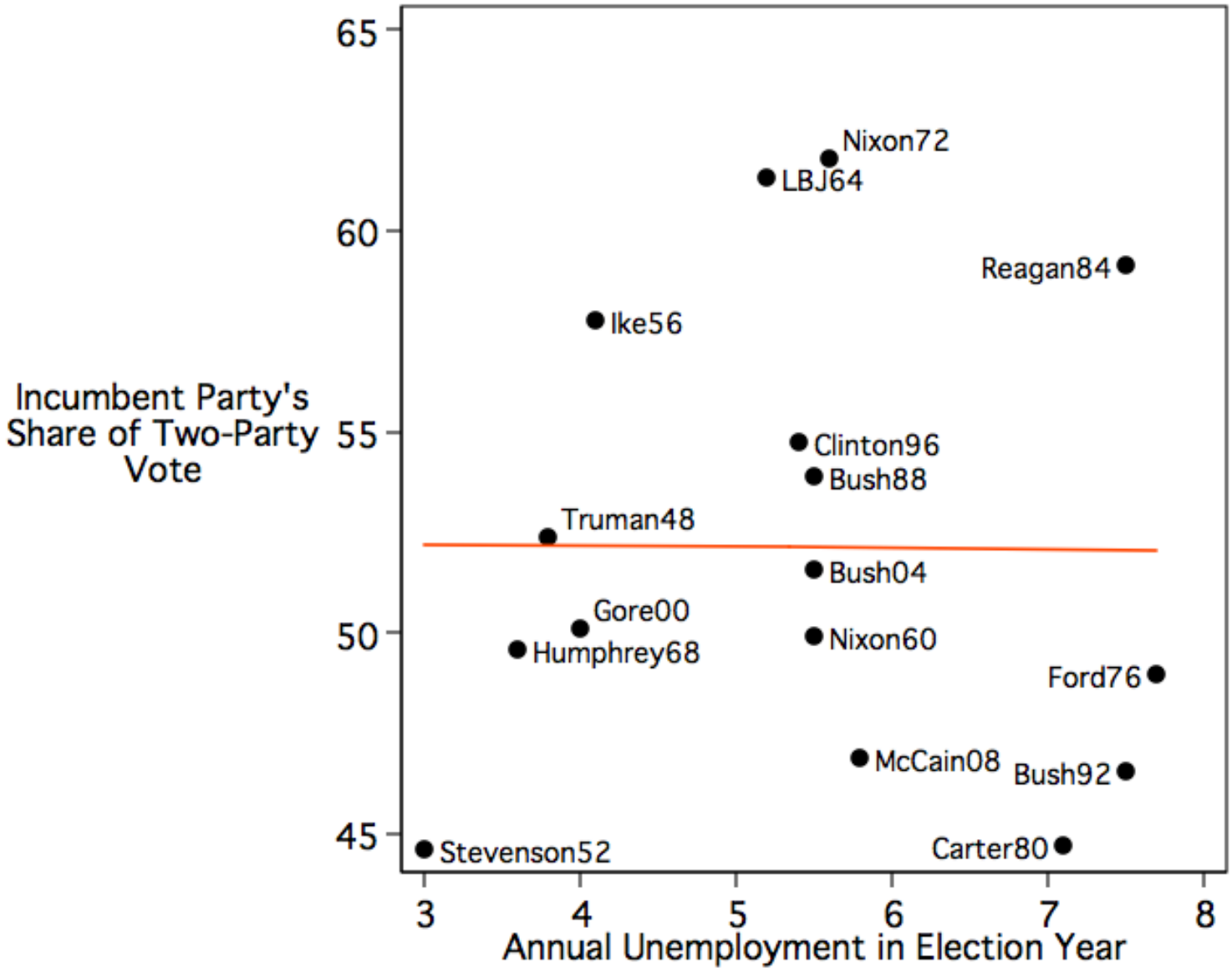
Scatterplots

What are we looking for?

- Form/pattern
- Direction
- Strength
- Outliers


Nationally, Trump got 2% more of the vote than Romney
Trump share of the two−party vote in 2016
Romney share of the two−party vote in 2012
0.3
0.4
0.5
0.6
0.7
WY
WV
ND
OK
ID
SD
KY
TN
MT
KS
UT
LA
AK
SC
TX
IA
OH
GA
AZ
NC
FL
VA
ME
NM
OR
DE
RI
IL
WA
NY
MD
VT
HI

(Gelman, 2016)


Incumbent Party's Share of Two-Party Vote
65
60
55
50
45
Nixon72
LBJ64
Reagan84
Ike56
Clinton96
Bush88
Truman48
Bush04
Gore00
Humphrey68
Nixon60
Ford76
McCain08
Bush92
Stevenson52
Carter80
3
4
5
6
7
8
Annual Unemployment in Election Year

(Masket 2011)

So how to we quantify this?

Correlation. But first . . .

Standardizing variables

$$\frac{x - \bar{x}}{s}$$

Example: Populations of New England states

	x	$\frac{x-\bar{x}}{s}$
CT	3.5m	0.48
ME	1.3m	-0.47
MA	6.6m	1.83
NH	1.3m	-0.47
RI	1.0m	-0.59
VT	0.6m	-0.78

$\bar{x} = 2.40 \quad s = 2.29$

Computation: Average of the products of the standardized values

$$r = \frac{1}{n-1} \sum \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right)$$

What does a positive correlation mean? Negative?

Visualizing correlation

Correlation $r = 0$

Correlation $r = -0.3$

Correlation $r = 0.5$

Correlation $r = -0.7$

Correlation $r = 0.9$

Correlation $r = -0.99$

Visualizing correlation

`http://guessthecorrelation.com/`

Facts about correlation

- Linear only
- **Not** causal
- Unit-free
- $-1 \leq r \leq 1$
- Sensitive to outliers

Regression: The big picture

What we want to do is the following:

- Assume we have two variables where the "outcome"" is interval(ish)
- Is there an "association" between them?
- Is it statistically significant (next class)?
- Estimate "expected values" for an outcome variable given a set of covariates

Some preliminaries

$Y =$ Response variable/ Dependent variable/
Outcome variable/Explained variable/ Left-hand side

Some preliminaries

Y = Response variable/ Dependent variable/
Outcome variable/Explained variable/ Left-hand side

X = Explanatory variable/ Independent variable/
Treatment Variable/ Right-hand side

How might Y and X be related?

Some preliminaries

Y = Response variable/ Dependent variable/
Outcome variable/Explained variable/ Left-hand side

X = Explanatory variable/ Independent variable/
Treatment Variable/ Right-hand side

How might Y and X be related? A line of course!

$$Y = \alpha + \beta X$$

Here α is the Y-intercept and β is the slope of the line.