PSC 400

SYRACUSE UNIVERSITY

DATA ANALYTICS FOR POLITICAL SCIENCE

LINEAR REGRESSION

COUNTRIES.CSV

variable	description
country	name of the country
gdp	country's GDP from 2005 to 2006 (in trillions of local currency units)
prior_gdp	country's GDP from 1992 to 1993 (in trillions of local currency units)
light	country's average level of night-time light emissions from 2005 to 2006 (in units on a scale from 0 to 63, where 0 is complete darkness and 63 is extremely bright light)
prior_light	country's average level of night-time light emissions from 1992 to 1993 (in units on a scale from 0 to 63, where 0 is complete darkness and 63 is extremely bright light)

EXAMPLE

Table 4.5. 2012 US Presidential Election Data.

Variable	*Description*
`state`	abbreviated name of the state
`Obama`	Obama's vote share (percentage)
`Romney`	Romney's vote share (percentage)
`EV`	number of Electoral College votes for the state

- pres12.csv
- How does Obama's vote share in 2012 depend on his 2008 vote share?

EXAMPLE

Table 4.1. 2008 US Presidential Election Data.

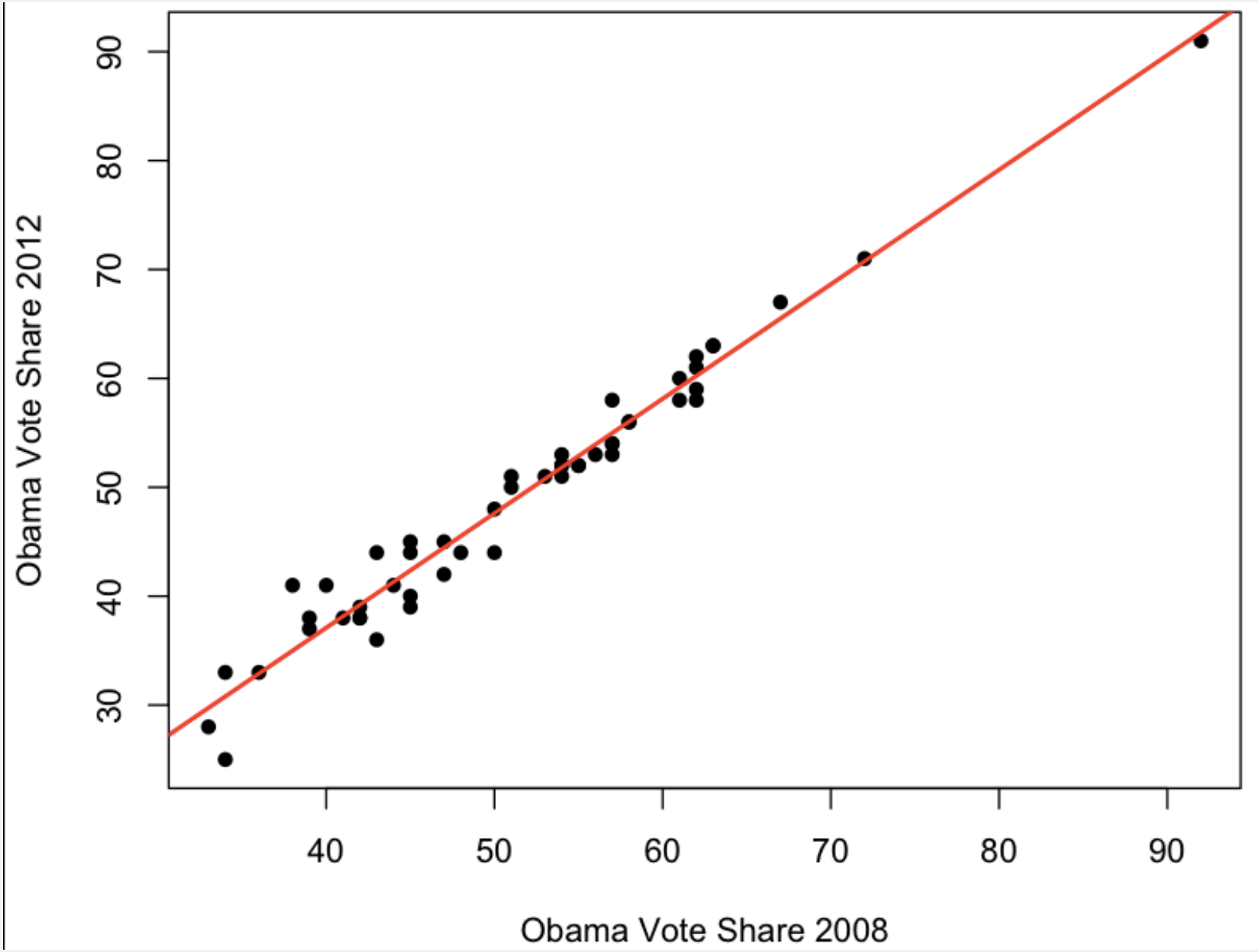
Variable	*Description*
`state`	abbreviated name of the state
`state.name`	unabbreviated name of the state
`Obama`	Obama's vote share (percentage)
`McCain`	McCain's vote share (percentage)
`EV`	number of Electoral College votes for the state

- pres08.csv

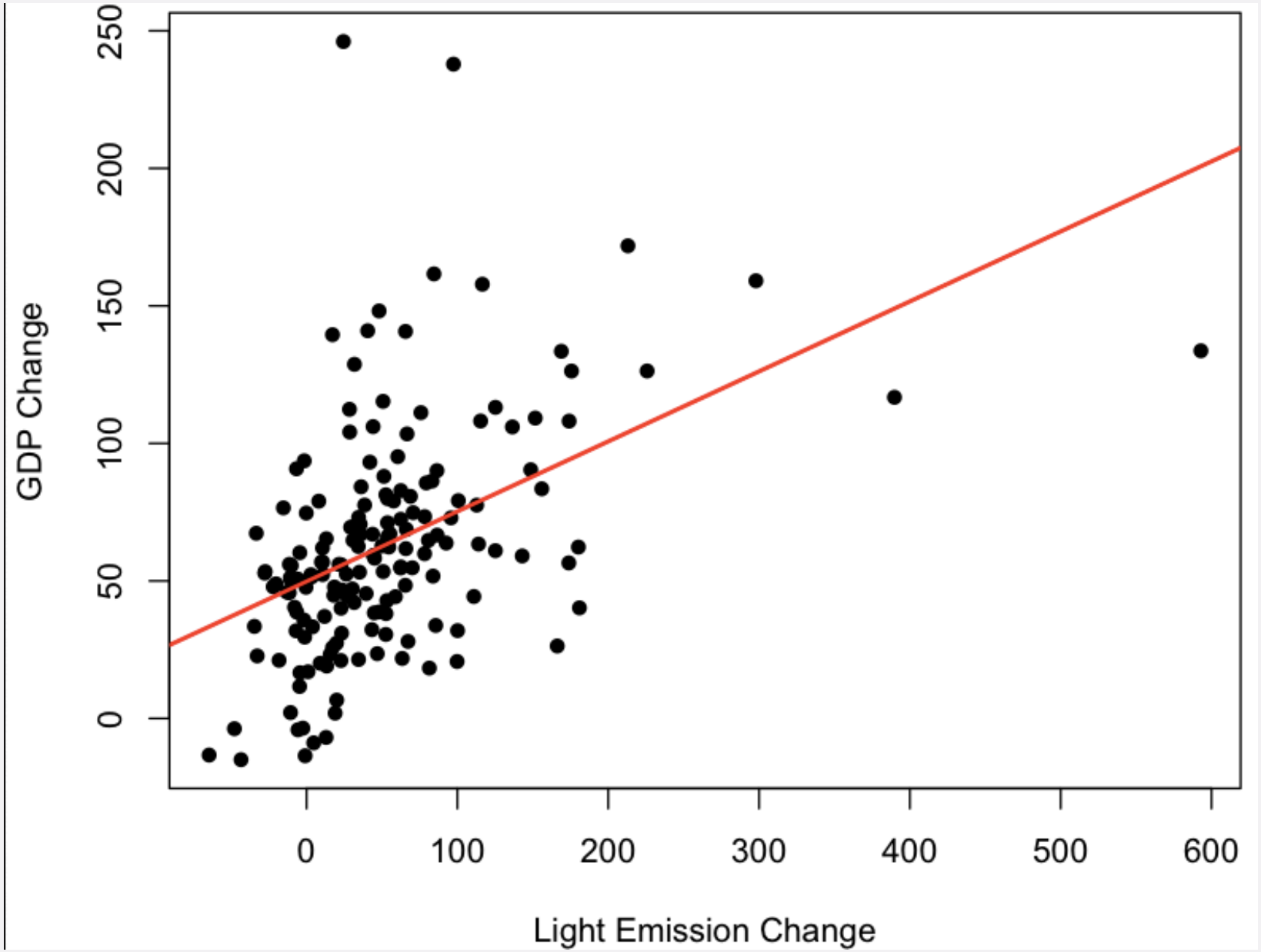
EXERCISE

- Create a scatterplot of Obama's 2008 vote share (x-axis) and 2012 vote share (y-axis)
- Estimate a linear regression to predict his 2012 vote share using his 2008 vote share
- Add the regression line to your scatterplot
- Interpret the intercept and slope coefficients
- What's the predicted 2012 vote share for a state where he got 50% in 2008?

COMPARE


Obama Vote Share 2012
30
40
50
60
70
80
90
40
50
60
70
80
90
Obama Vote Share 2008

COMPARE


GDP Change
0
50
100
150
200
250
0
100
200
300
400
500
600
Light Emission Change

EXPLANATORY POWER MEASURE

- **Need: measure of how well independent variable explains dependent variable in a linear regression**

EXPLANATORY POWER MEASURE

- Measure is called R^2
- R^2 tells us how much variation of the dependent variable is explained by the independent variable
  - Between 0 and 1
  - 0: The independent variable explains *none* of the variation in the dependent variable
  - 1: The independent variable explains *all* of the variation in the dependent variable

EXERCISE

- Quality of Government data
- DV: Corruption perceptions index: ti_cpi
- IV: pick a (numerical) variable
- Scatterplot, linear regression line, R-squared