Course Materials May Not Be Distributed or Posted Electronically

These course materials are the sole property of Dr. Todd M. Gross. They are strictly for use by students enrolled in a course taught by Dr. Gross. They may not be altered, excerpted, distributed, or posted to any website or other document-sharing service.

© 2010–2017 Todd M. Gross

PSTAT 126
Regression Analysis

Dr. Todd Gross

Department of Statistics and Applied Probability

University of California, Santa Barbara

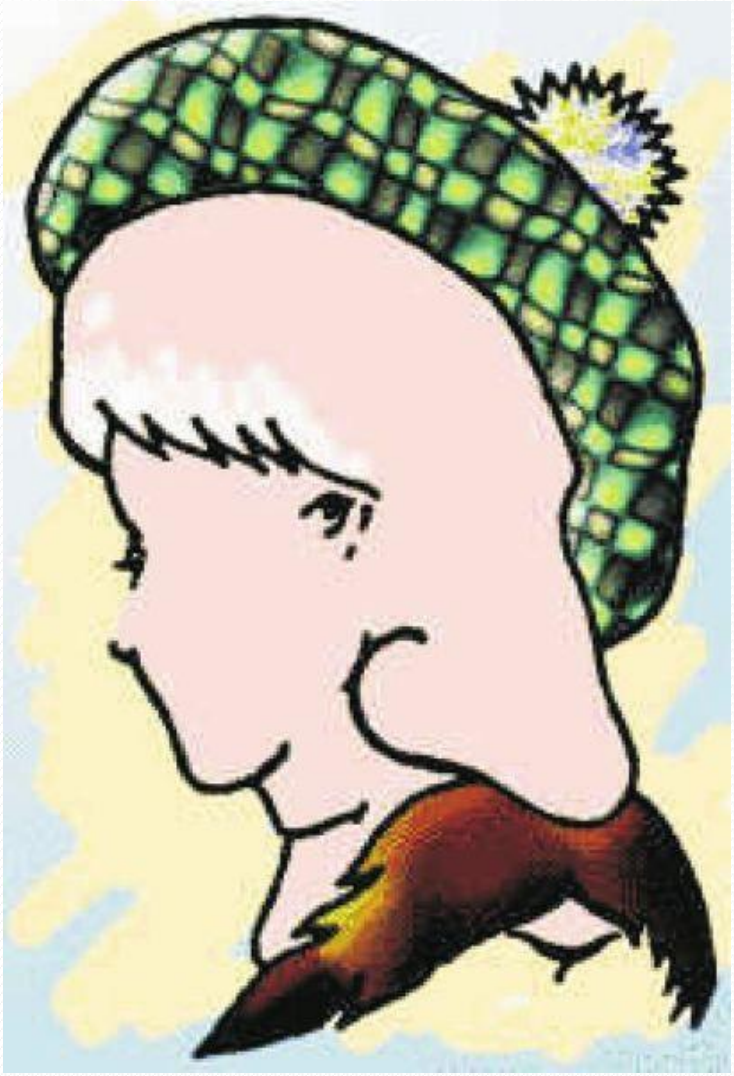
Lecture 1

Lecture Outline

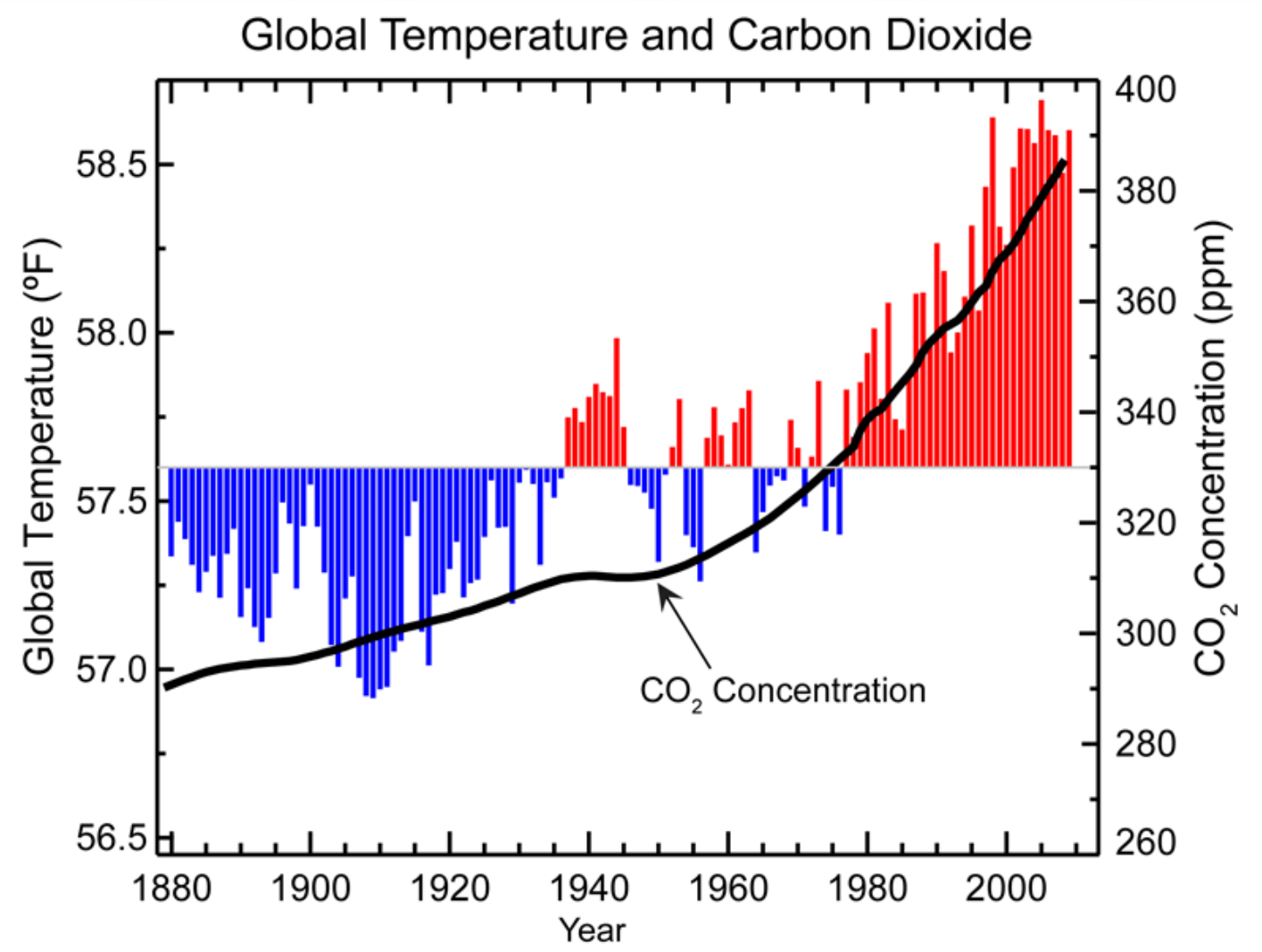
- What is Regression Good For?
- Administration/Syllabus
- How To Do Well in this Course
- My Background
- Review of Basic Statistics

What is Regression Good For?

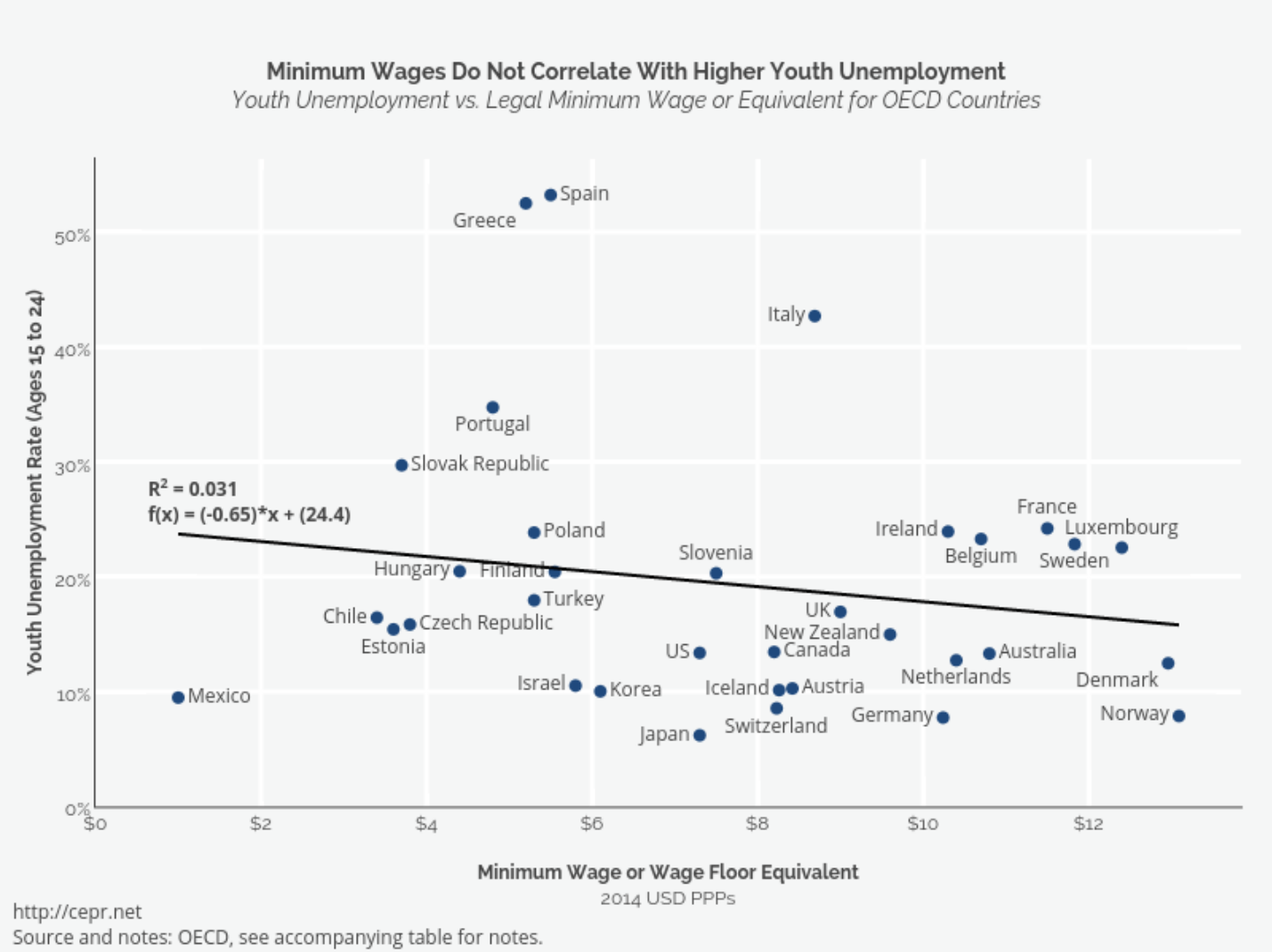
What Do You See?



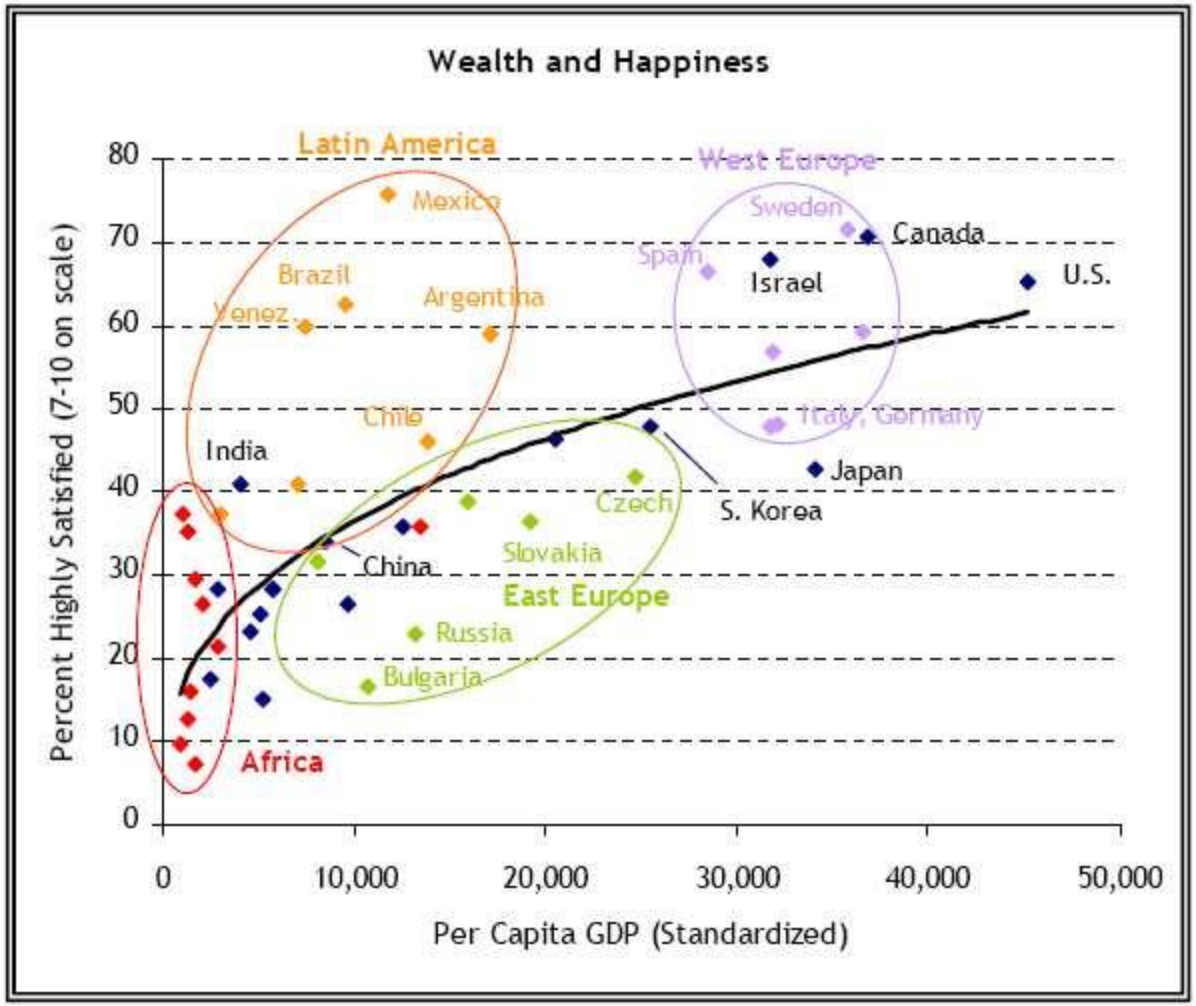
Does Human Activity Affect Climate?



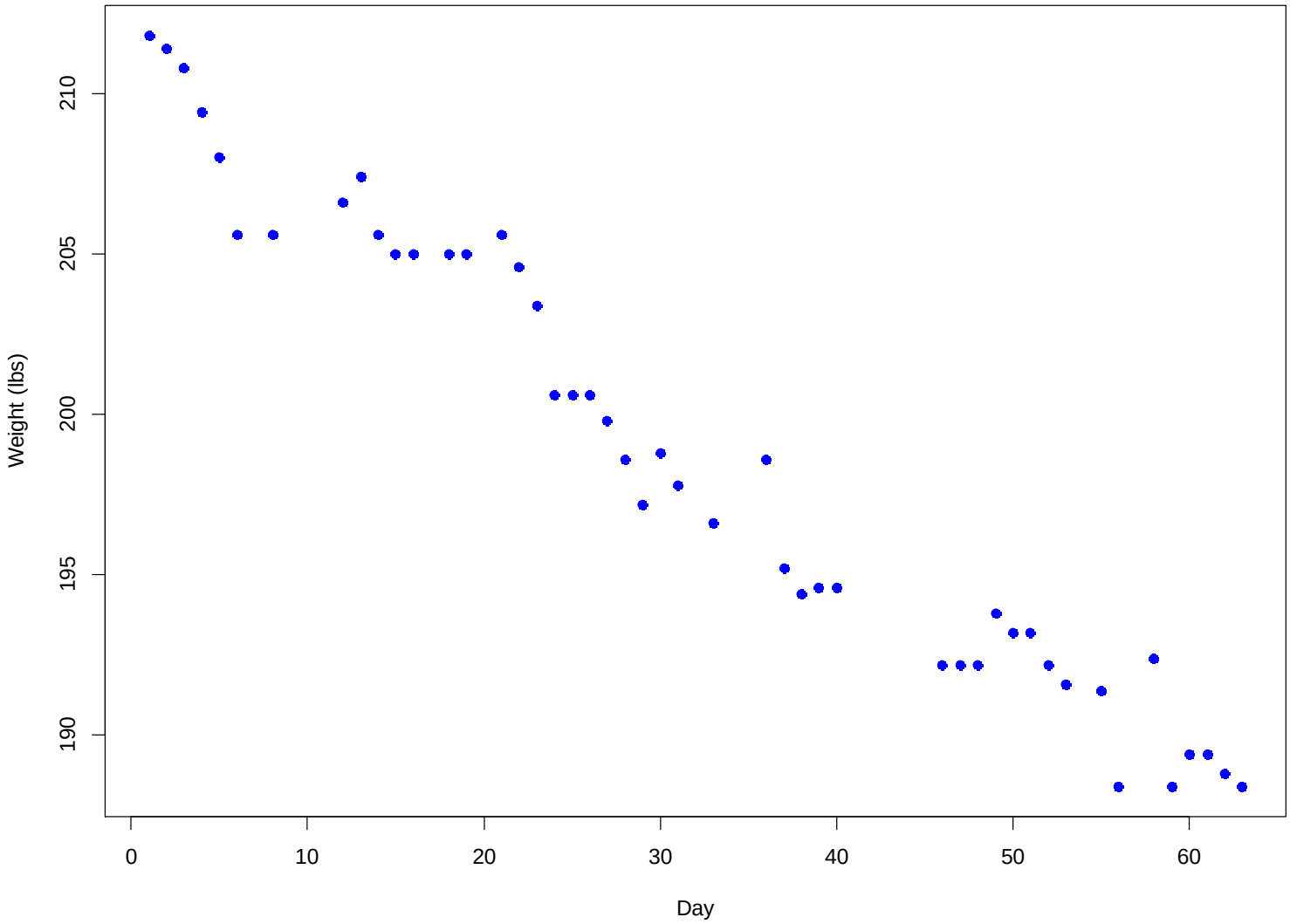
Does Higher Minimum Wage Lead to Higher Unemployment?



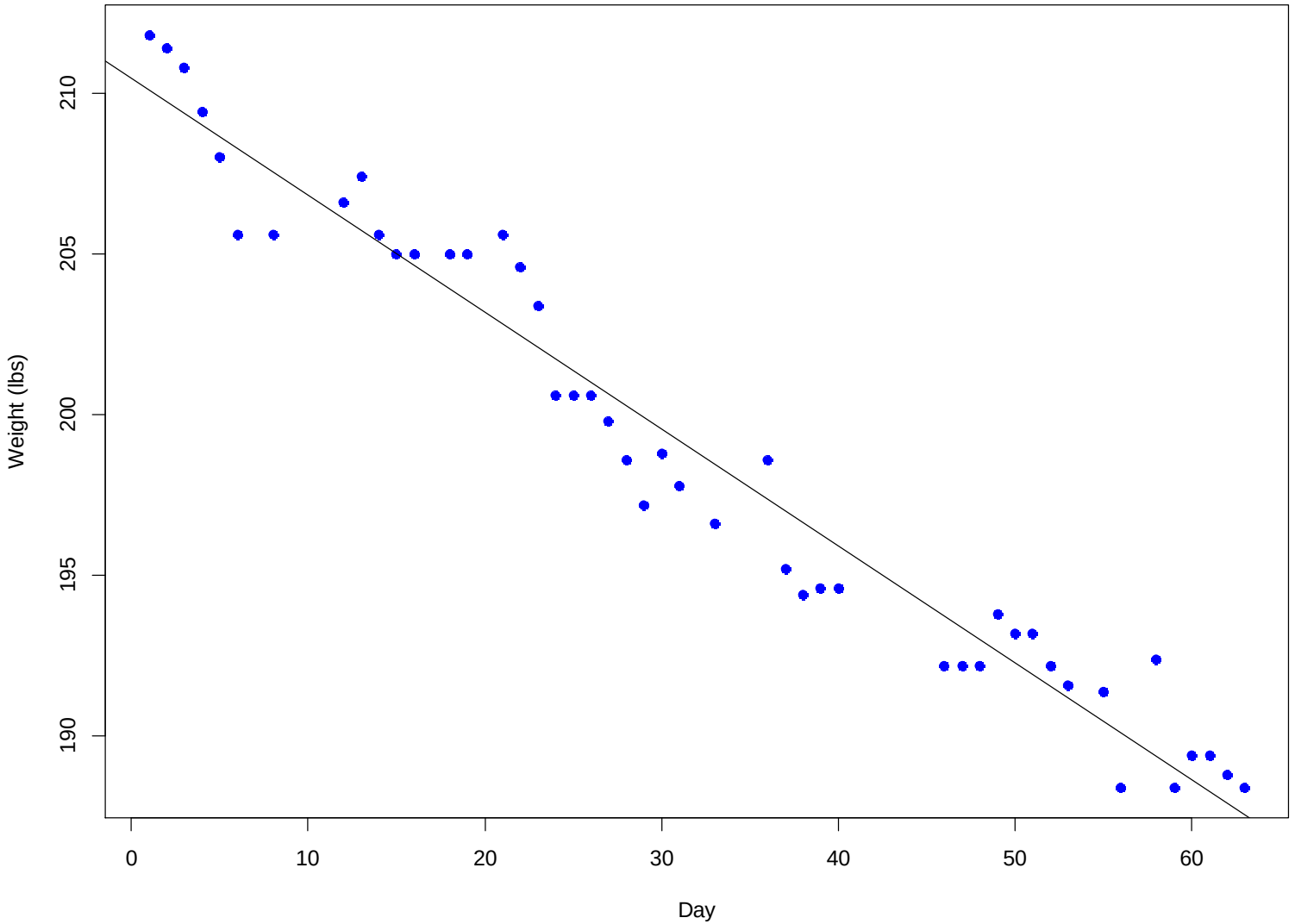
Can Money Buy Happiness?



Weight Loss Example - Scatterplot



Weight Loss Example – Regression Line



Administration & Syllabus Review

Course Materials

Lecture

- The lectures will focus on a conceptual understanding of linear regression
  - Slides will support the lecture, but you will need to take notes.

Labs

- The labs will provide instruction on how to use R

Textbooks

- *Linear Models with R,* Second Edition by Julian J. Faraway.
  - This text is **optional**.
  - Focus on using R to analyze linear regression models
- *An Introduction to Statistical Learning* by Gareth James, Trevor Hastie and Robert Tibshirani.
  - **DO NOT PURCHASE**
  - May be downloaded at http://www-bcf.usc.edu/~gareth/ISL/.

Homework

- Homework will be assigned at the beginning of each week
- Due in your enrolled section the following week
- Give complete answers and show work
- You will be writing formulas and annotating R output, so you will need to turn in hard copies (cannot be submitted by email)
- Please make your printouts compact (reduce size of figures from R)
- Late homework will NOT be accepted

Grading

Grading Component	Percentage	Notes
Attendance	12%	Lecture and Lab, No Excused Absences
Participation (Lecture)	3%	Based on participation in lecture & Dr. Gross' office hours
Homework	15%	Due in Enrolled Lab
Class Project	15%	Work in Teams of 2, Due in Last Lab
Exam #1	25%	Thursday, Nov 2 (in lecture)
Exam #2	30%	Thursday, Dec 7 (in lecture)

Active Participation Key to Doing Well

- Keep up with reading and assignments
- Active participation produces mastery of material
- Asks questions – don't get left behind
- Answer questions – test you knowledge, active association
- Study with others
- Relate this material to what you already know (or are learning in other courses)

My Background

- PhD in Cognitive Psychology (UCLA)
- Practicing Statistician for 35 years
- Career as Head of Biostatistics in Biotech Industry (Allergan, Medtronic, Boston Scientific, Kythera)
- Scientific Consultant for 3D Communications (FDA Advisory Committee support)

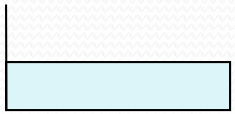
Review of Basic Statistics

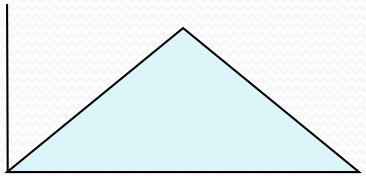
3 Key Attributes of a Set of Values

- What are the three things we need to describe a set of values (i.e., a distribution)?
  - Shape
  - Center
  - Spread

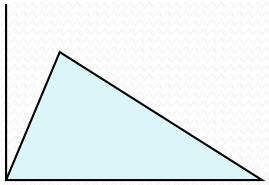
Shape - Types of Distributions

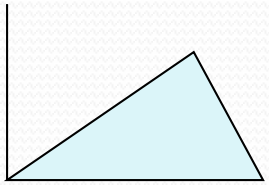
- Uniform



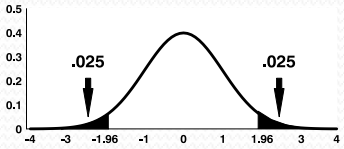


- Symmetric
- Skewed
  - Positive or negative

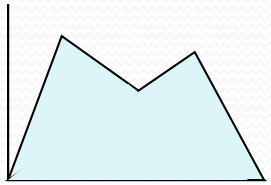




- Normal



- Bimodal



Measures of Center

- Mode
  - Most frequent score
  - Sensitive to only a few scores in distribution
- Median
  - Middle score
  - 50% of distribution above, 50% below
- Mean
  - Average score
  - Sensitive to all scores in distribution

Measures of Spread (Dispersion)

- Range = highest score – lowest score
  - Sensitive to only two scores in the data set
- Variance
  - Increases as a function of variability
- Standard Deviation
  - The “average” or “typical” deviation

Which Set of Scores Has More Variability?

- {1,2,3,4,5} or {6,7,8,9,10}
- {11,12,13,14,15} or {9,11,13,15,17}
- {1,2,3,4,5} or {1,2,3,4,25}

Variance and Standard Deviation

	X	Mean(X)	X - Mean(x)	$(X - Mean(x))^2$
	1	3	-2	4
	2	3	-1	1
	3	3	0	0
	4	3	1	1
	5	3	2	4
Sum	15		0	10
Average	3		0	2

- Variance(X) = 2
- Standard Deviation(X) = Sqrt(Var) = Sqrt(2) = 1.414

Mean versus Standard Deviation

Data = 1,2,3,4,5 Mean = 3 SD = 1.4	Data = 2,4,6,8,10 Mean = 6 SD = 2.8
Data = 4,5,6,7,8 Mean = ? SD = ?	Data = 10,20,30,40,50 Mean = ? SD = ?

Variance is the Key to Regression

- Variance captures all of the interesting things that happens in Life!
  - Imagine a world where everyone and everything was the same
- Variability is what allows us to investigate the relationship between two variables
  - If either of the variables has little or no variance, then it is hard to see a relationship
- Regression provides tools for “explaining” variance
  - Example: Can we “explain” your score on the midterm by knowing other variables?