

# Lessons Learned in Transitioning from *Intro to Statistics* to *Reasoning with Data*

Rebecca Nugent

Carnegie Mellon University Statistics

August 1, 2017

# Carnegie Mellon University

- ▶ Private university in Pittsburgh, PA
- ▶ R1 research university designation
- ▶  $\approx$  6500 undergrads, 6500 grads
- ▶ Six undergraduate colleges (admission is college-specific)
  - ▶ College of Fine Arts
  - ▶ Dietrich College of Humanities & Social Sciences (including Information Systems)
  - ▶ College of Engineering
  - ▶ Mellon College of Science
  - ▶ School of Computer Science
  - ▶ Tepper School of Business

# Carnegie Mellon Statistics

- ▶ Dietrich College of Humanities & Social Sciences (1200)
  - ▶ Economics (joint in Tepper)
  - ▶ English
  - ▶ History
  - ▶ Information Systems
  - ▶ Institute for Politics and Strategy
  - ▶ Modern Languages
  - ▶ Philosophy
  - ▶ Psychology
  - ▶ Social and Decision Sciences
  - ▶ Statistics
- ▶ Undergraduate Statistics: around 450 primary/add'l majors
  - ▶ Statistics (Concentration: Open, Math, Neuroscience)
  - ▶ Economics-Statistics
  - ▶ Statistics and Machine Learning

# Dietrich General Education Curriculum

## *Current:*

- ▶ Computing @CMU
- ▶ Introduction to Statistics
- ▶ Writing/Interpretation
- ▶ History survey course
- ▶ Freshmen seminar
- ▶ Remaining courses are selected from categories:  
Communicating, Creating, Deciding, Modeling, Reflecting

Students largely feel like the courses aren't well-connected to each other; hard to get into some of them; often end up just looking for one to take or trying to get substitutions

Courses often aren't designed for general education

# Dietrich General Education Curriculum

In midst of multiple-year revamp/design of new requirements

College-wide committees looking for some general consensus  
(you can imagine how well that's going)

*Proposed for Fall 2018:*

- ▶ Reasoning with Data
- ▶ Writing minis: choose 2 of 3 themes (e.g. Writing with Data)
- ▶ Freshmen and Junior seminars
- ▶ More interdisciplinary courses across departments
- ▶ Experiential Learning and Service
- ▶ Portfolio and Self-Reflection
- ▶ Broader focus on courses that are truly general education

So how do we develop a modern general education course in statistics/data science for liberal arts and *some* STEM that

1. the students don't hate
2. the students find useful
3. doesn't really require any kind of computing (background)
4. illustrates the richness/diversity of types of data and interdisciplinary problems in data science
5. is reproducible and can grow/build on itself

# Where we're starting

## *36-201: Statistical Reasoning and Practice*

- ▶ Between 120-240 each semester; required for all Dietrich
- ▶ Standard class structure: lecture, HWs, labs, exams
- ▶ Lab: hands-on practice of concepts
  - ▶ using Minitab as the software package
  - ▶ very results-oriented (calculate, basic interpretation)
  - ▶ staffed by master's/undergraduates (primarily) in Statistics
- ▶ Concepts: EDA, experimental design, elem. probability, discrete distributions, normal distribution, sampling distributions, confidence intervals, significance tests, one-way ANOVA, contingency tables, linear regression
- ▶ Emphasis is on learning steps of analysis, correct formulas to use for which scenario (“cookbook” approach), basics

# Where we're starting

## *36-201: Statistical Reasoning and Practice Challenges:*

- ▶ Can be viewed as “dry”  
(common complaint about intro stat courses in general)
- ▶ Formula-driven, learning a set of steps
- ▶ Hard to have students see the whole picture
- ▶ Primarily focused on continuous and categorical data
- ▶ Students use different software packages in different disciplines
- ▶ Intro examples tend to be more STEM-driven, social science applications, very little humanities
- ▶ Contains almost no modern material or applications



# What we're doing now

Interviews w/ different Dietrich depts about issues/needs

- ▶ Students don't know the concepts
- ▶ Get tied to the specific software steps
- ▶ Can't see the big picture
- ▶ Class isn't really for them

*Our goals:*

- ▶ Modernize the course
- ▶ Emphasize concepts; tell stories with data
- ▶ More student-driven inquiry
- ▶ More adaptive material
- ▶ learn how different students interact with data
- ▶ **Convert the class into a community**

# What we're doing now

## *Course Vision*

- ▶ Concepts, concepts, concepts
- ▶ Interpretation, interpretation, interpretation
- ▶ Remove computing cognitive load completely
- ▶ Interact/engage with problems
- ▶ Data: Continuous, Categorical, Text, Images, Networks, Time Series (neuroscience, economics, etc)
- ▶ Use atypical data types to illustrate intro level concepts
- ▶ Include overview of modern statistics methodology taught at conceptual level (classifiers, clustering, “big data”...)

# What we're doing now

## *Class Structure*

- ▶ Lectures, weekly HW
- ▶ “Discussion Labs”
  - ▶ small (15-20)
  - ▶ staffed by TAs from Stat and other Dietrich units
  - ▶ presented as a case study/story
  - ▶ combination of questions and interpretation discussion  
What do these results mean? What might we do next?
- ▶ Quizzes/Final Exam
- ▶ Data Story project(s)

# What we're doing now

## *Interactive Software Platform (ISLE)*

- ▶ No coding. More direct interaction with the concepts
- ▶ Easy to adapt new hypotheses, problems during class/lab
- ▶ Can collect answers from students; propagate through labs
- ▶ "Data Set Explorer": upload (formatted) data, variables
- ▶ Students can save graphs and work to editors that create websites/documents for a portfolio
- ▶ We can collect information on clicks, decisions, times, etc - how do students explore data?
- ▶ Combining tools like Java Script with R
- ▶ Built in modular form; can "mix-and-match"

Philipp Burckhardt, Tues 8/1, 11:05am, Session #344, CC-309

# What we're doing now

*Let's look at some examples:*

Interactive Venn Diagram

Conditional Probability with Text

Collecting Data

Forest Fires

Everything they do, everywhere they go - we can track them.  
I promise it's not creepy.

# What we're doing now

## *Data Story Projects:*

- ▶ Describe/Examples
- ▶ Case Study Projects
- ▶ Students can upload/contribute data sets to the explorer (“normal”, text, images, etc)
- ▶ Projects can be suggested or self-defined
- ▶ Building a repository/community that grows with instructor supervision

# What we're doing now

## *Data@Dietrich*

- ▶ Video series from across Dietrich - What is Data to you?
- ▶ Assigned to students as HW or used for case studies
- ▶ **Example from Modern Languages**
- ▶ Students get to see breadth/depth of problems
- ▶ Statistics majors see other kinds of data
- ▶ Non-stat majors see potential future projects
- ▶ Faculty get advertising for all their projects and classes at minimal cost to them
- ▶ Everybody gets an active public-facing website repository

# What we're going to do this year

*How do liberal arts and Stat students interact with data?*

- ▶ Track
- ▶ Different pathways
- ▶ Adapt as the lab is happening
- ▶ Different discussion labs for different groups?



# What we're going to do this year

## *Discover Data@Dietrich*

- ▶ Currently have Tartan Data Science Cup
- ▶ Introductory level workshops for college
- ▶ Joint between Statistics and Dietrich unit
- ▶ Present problem and question
- ▶ Data, beginning analysis
- ▶ Workshop the problem, see what happens

# Takeaways

- ▶ Building Data Science courses is not enough
  - ▶ Humanities/Liberal Arts students may need something completely different
  - ▶ Need to figure out what that is
- 
- ▶ Convert class to community
  - ▶ Give “ownership” to stakeholders
  - ▶ Community should grow, evolve with contributions
  - ▶ Build easily-traversed bridges

# The Team/Acknowledgments