Lessons Learned from Running a Distance-Based Course in Data Visualization

Cramér Society Fall Meeting 2021

Johan Larsson

Department of Statistics, Lund University

October 25, 2021

LUND
UNIVERSITY

Overview

Course Design

Course Content

Lessons Learned

Course Design

Motivation

- Data visualization is an integral part of statistics
  - Exploring data
  - Understanding theory
  - Diagnosing models
  - Presenting results
- Useful in many other professions

Aims

Overall Aim

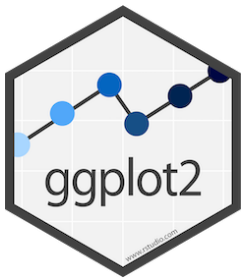
Teach data visualization using modern tools, focusing on both theory and practice but with emphasis on the former.

Design-Specific Specific Aims

- Focus teacher resources on feedback and guidance
- Allow asynchronous progress
- Enable the course to scale well to many students
- Use free (but modern) tools and literature
- Encourage student interaction

Format

- 4 ECTS credits ($\approx$ 10 hours/week)
- Entirely distance-based
- Based on R and the R-package ggplot2
- Platform: Canvas
- Slack workspace for discussions
- Zoom for workshops
- Freely available literature and tools



Course Content

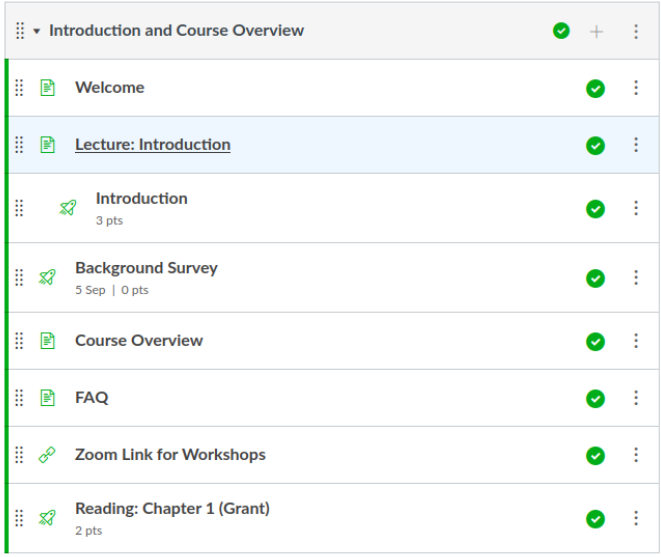
Course Components

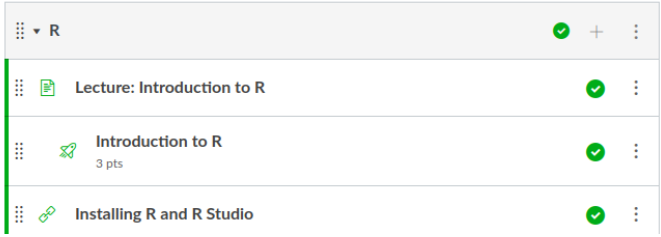
Course Activities

- Pre-Recorded Lectures
- Reading Assignments
- Worked Examples
- Online Workshops

Examination

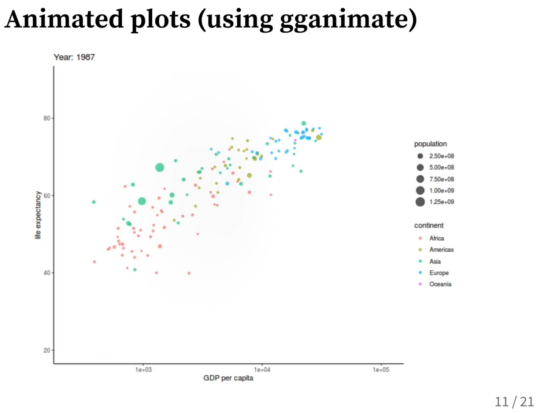
- Quizzes
- Assignments
- Project





Lectures

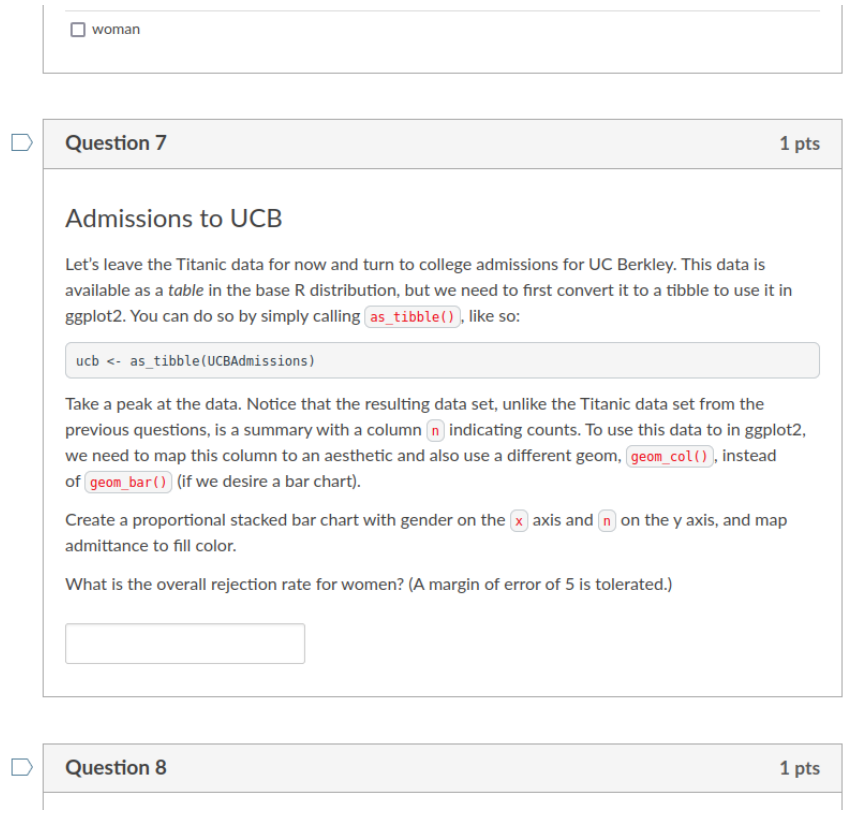
- Pre-recorded
- Short (5–10 minutes)
- Scripted
- High production values



11 / 21

Quizzes

- Three types
  - Lecture quizzes
  - Reading quizzes
  - Practice quizzes
- Mostly automatically graded



☐ woman

Question 7 1 pts

Admissions to UCB

Let's leave the Titanic data for now and turn to college admissions for UC Berkley. This data is available as a *table* in the base R distribution, but we need to first convert it to a tibble to use it in ggplot2. You can do so by simply calling `as_tibble()`, like so:

```
ucb <- as_tibble(UCBAdmissions)
```

Take a peak at the data. Notice that the resulting data set, unlike the Titanic data set from the previous questions, is a summary with a column `n` indicating counts. To use this data to in ggplot2, we need to map this column to an aesthetic and also use a different geom, `geom_col()`, instead of `geom_bar()` (if we desire a bar chart).

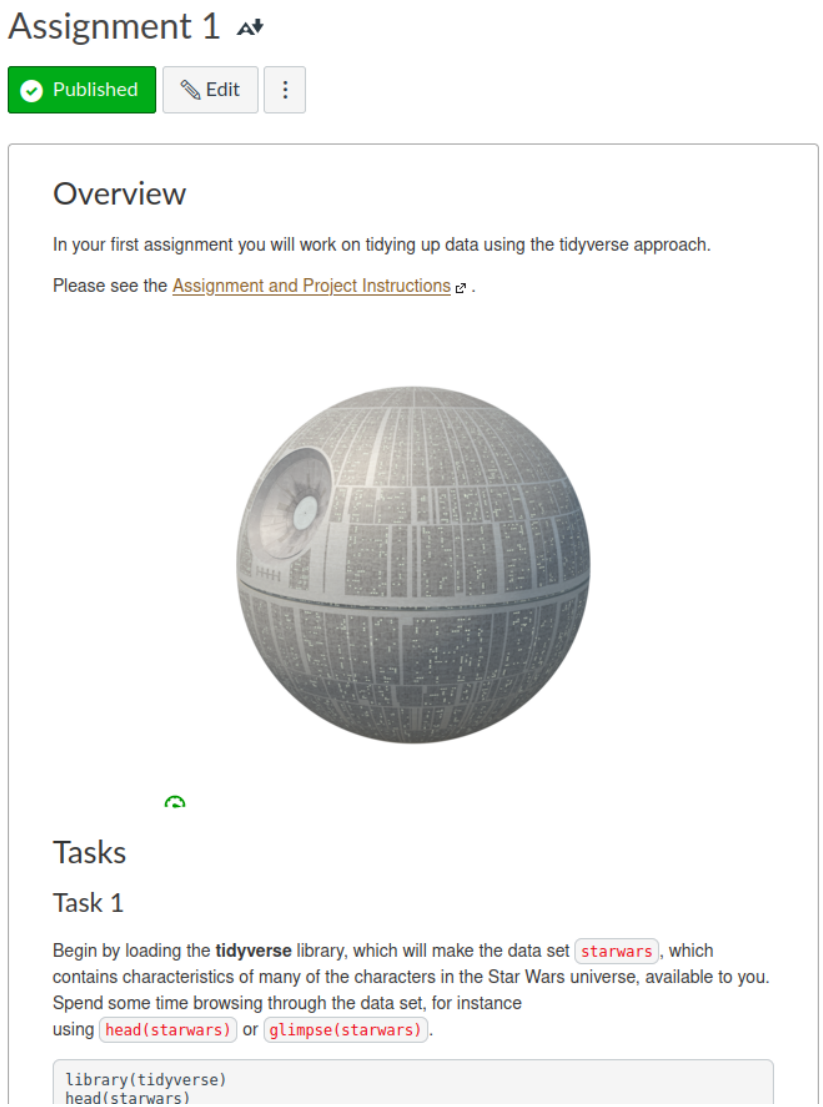
Create a proportional stacked bar chart with gender on the `x` axis and `n` on the y axis, and map admittance to fill color.

What is the overall rejection rate for women? (A margin of error of 5 is tolerated.)

Question 8 1 pts

Assignments

- Three tasks in each
- Progressively freer (and more challenging)
- Peer feedback
- Short feedback loops
- Encourage reproducible submissions (R Markdown)



Project

- Free choice of "research question"
- Free choice of data set (under certain restrictions)
- Large variation in results

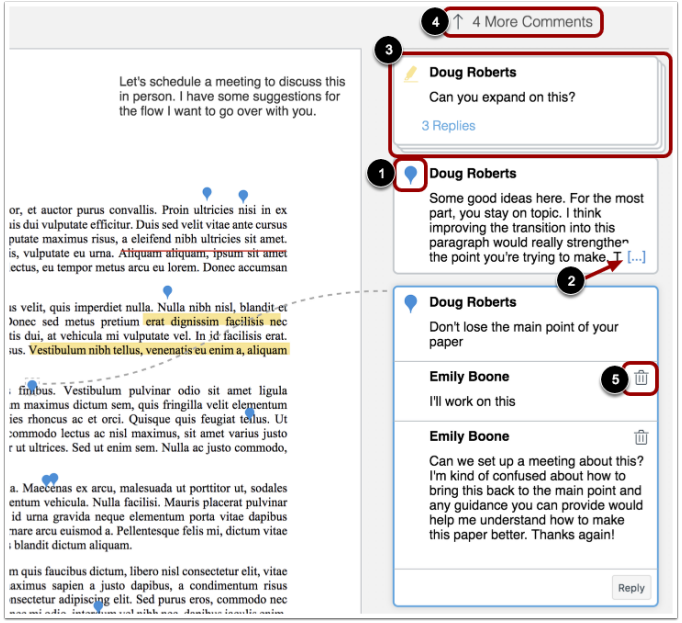
Lessons Learned

Pre-Recorded Lectures: Costly but Worthwhile

- Excellent student response
- Frees up time for real interaction
- Large time investment up-front
- Best for old (stable) courses

Peer Feedback: Great Value

- Saves time
- Fast feedback
- From student perspective works best when assignments are free. Opposite situation from teacher perspective.
- Mixed engagement



Quizzes: Hard to Design, but Useful

- Hard to design automatically graded quizzes (at least for data visualization)
- Easy to lose track of students.
  - Solution: use a few manually graded tasks
- One second attempt seems appropriate, but leads to inflation
- "Encourages" students to actually read course literature

Using R: Feasible but not Frictionless

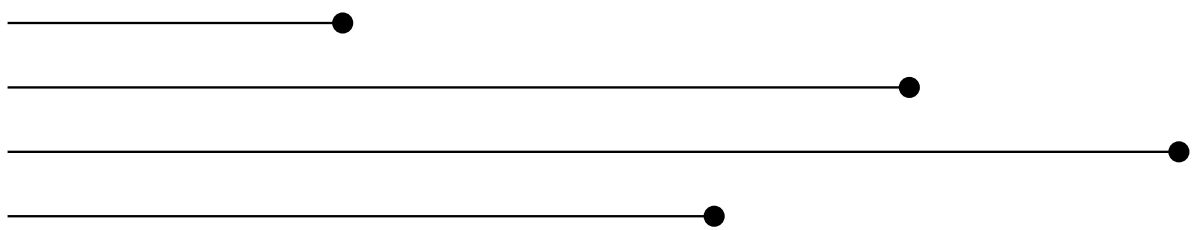
- Anyone can learn how to use R
- Makes assisting students easier
- Speeds up revisions
- Needs introduction
- Divides students into two groups

Retention

- Register many students! $\approx 50\%$ drop out.
-

Asynchronicity: A Mixed Bag

- Great for some students
- Does not combine so well with peer (and general) feedback



Conclusions

- Works well overall
- Peer reviews are great
- Quizzes are efficient in stimulating learning
- Automatic grading is useful but hard to design well