

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

| Introduction and Course Overview | | | ✓ | + | ⋮ |
|----------------------------------|----------------------------|---------------|---|---|---|
| 📄 | Welcome | | ✓ | | ⋮ |
| 📄 | Lecture: Introduction | | ✓ | | ⋮ |
| 📄 | Introduction | 3 pts | ✓ | | ⋮ |
| 📄 | Background Survey | 3 Sep 0 pts | ✓ | | ⋮ |
| 📄 | Course Overview | | ✓ | | ⋮ |
| 📄 | FAQ | | ✓ | | ⋮ |
| 🔗 | Zoom Link for Workshops | | ✓ | | ⋮ |
| 📄 | Reading: Chapter 1 (Grant) | 2 pts | ✓ | | ⋮ |

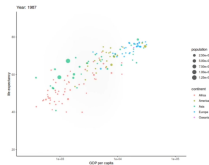
| R | | | ✓ | + | ⋮ |
|---|----------------------------|-------|---|---|---|
| 📄 | Lecture: Introduction to R | | ✓ | | ⋮ |
| 📄 | Introduction to R | 3 pts | ✓ | | ⋮ |
| 🔗 | Installing R and R Studio | | ✓ | | ⋮ |

Lectures

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



Animated plots (using gganimate)



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 Berkeley. 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)

Assignment 1



Overview

In your first assignment you will work on tidying up data using the tidyverse approach.

Please see the [Assignment and Project Instructions](#).



Tasks

Task 1

Begin by loading the **tidyverse** library, which will make the data set `starwars`, which contains characteristics of many of the characters in the Star Wars universe, available to you. Spend some time browsing through the data set, for instance using `head(starwars)` or `glimpse(starwars)`.

```
library(tidyverse)
head(starwars)
```

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

The screenshot displays a peer feedback interface. On the left, a document with placeholder text is shown. On the right, a sidebar contains comment threads. The top thread, by Doug Roberts, asks 'Can you expand on this?' and has 3 replies. The second thread, also by Doug Roberts, points to a specific paragraph and suggests improvements. The third thread, by Emily Boone, asks 'Can we set up a meeting about this?'. The interface includes a '4 More Comments' link at the top of the sidebar and a 'Reply' button at the bottom.

Let's schedule a meeting to discuss this in person. I have some suggestions for the flow I want to go over with you.

et, et sector parva corvalla. Prois aleticus nati in ex de du vulgare efficitur. Duis sed velit vitae ante curas puate maxime riuu, a delfend nbi ultrices vii amet. a, vulgare ex urna. Aliquam aliquam: ipsum ut amet actus, ex tempus metus eros ex lorem. Donec accumsan

u velit, quis imperit nulla. Nulla nbi rid. Maecis et bene: sed natus pretius dui dignum facilis acis dat, at vehicula mi vulgare veli. In id facilis erat. ut. Vestibulum nbi nulla, venenatis eu eros a, aliquam

hinc. Vestibulum pulvinar odio sit amet ligula in maxime dictum seni, quis fringilla veli elegerant leo thecus ac et orci. Quisque quis fringilla illas. Ut, semmodo lectus ac nisl maxime, vii amet varius justo r ut ultrices. Sed ut eros seni. Nulla ac justo conmodo,

a. Maecenas ex eros, malesuada ut portitor ut, sodales etiam vehicula. Nulla facilis. Maecis placerat pulvinar id ante gravida neque elementum puru vitae dapibus vari eros eueurol a. Pellentesque feli mi, dictum vitae /hacit dictum aliquam.

in quis faucibus dictum, libero nisl consectetur elit, vitae accums nuptes a justo dapibus, a condimentum riuu suscipit adipiscing elit. Sed parva eros, conmodo nec

4 More Comments

Doug Roberts
Can you expand on this?
3 Replies

Doug Roberts
Some good ideas here. For the most part, you stay on topic. I think improving the transition into this paragraph would really strengthen the point you're trying to make. T [-]

Doug Roberts
Don't lose the main point of your paper

Emily Boone
I'll work on this

Emily Boone
Can we set up a meeting about this?
I'm kind of confused about how to bring this back to the main point and any guidance you can provide would help me understand how to make the paper better. Thanks again!

Reply

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

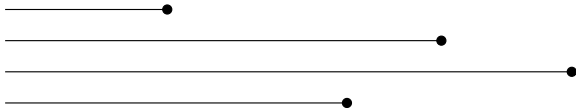
- Anyone can learn how to use R
- Makes assisting students easier
- Speeds up revisions
- Needs introduction
- Divides students into two groups

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