

# **Lessons Learned from Running A Distance-Based Course in Data Visualization**

**Cramér Society Fall Meeting 2021**

Johan Larsson

Department of Statistics, Lund University

October 27, 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 credits, half semester ( $\approx 10$  hours/week)
- Distance-based
- Uses 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

# Workshops

- 1-hour online sessions on Zoom, once per week
- Instruction and help with course content



# 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 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

- Short
- Progressively freer (and more challenging)
- Peer feedback
- Reproducible submissions (R Markdown) are encouraged

## 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
- Play, pause, slow down/speed up
- Large time investment up-front
- Lock-in effects

# 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 is shown with several paragraphs of placeholder text. Blue speech bubble icons indicate where comments have been made. On the right, a sidebar contains a list of comments. The top comment is from Doug Roberts, asking 'Can you expand on this?' with 3 replies. Below it is another comment from Doug Roberts, asking 'Don't lose the main point of your paper', with 5 replies. The interface includes a '4 More Comments' link at the top of the sidebar and a 'Reply' button at the bottom.

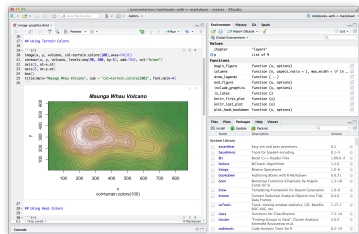


## Quizzes: Hard to Design, but Useful

- 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
- Gamifies the course (collect points)

# Using R and R Markdown: Feasible but not Frictionless

- Anyone can learn how to use R
- Makes assisting students easier
- Speeds up revisions
- Needs introduction
- Student background matters
- R Markdown is fantastic for assignments

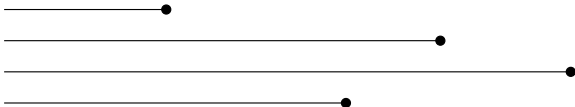


## Workshops: Great (Provided Someone Shows Up)

- Works great
- Remote control and sharing is effective
- Flexible (drop-in, drop-out)
- Problem is that few people show up

## Asynchronicity: A Mixed Bag

- Great for some students
- Does not combine so well with peer feedback
- Access to everything at once may not suit everyone



# Conclusions

- Pre-recorded lectures: high costs up-front, but large rewards down the road
- Peer reviews are great
- Quizzes are useful, but making well-made automatically graded quizzes is hard
- R Markdown is fantastic for student assignments
- Expect large variation in ambition and engagement
- Expect high drop-out rates