## Template

**Title UK:** Social Data Science

**Title for the IT-system STADS:** Social Data Science

**Financial course:** No

**Academic Level:** BA 3. year

**Semester placement:** Autumn

**Maximum of participants:** none

**Teacher**

* **Name:** Sebastian Barfort
* **CPHU Username:** jcw920
* **Use of guest speakers:** Yes
* **Use of teaching assistants:** ?

**Content:**  
The objective of this course is to learn how to gather and work with modern social science data. Increasingly, social data--data that capture how people behave and interact with each other--is available online in new, challenging forms and formats. This opens up the possibility of gathering large amounts of interesting data, provided that the analyst has the sufficient computer literacy. Consequently, being an effective social scientist means spending large fractions of our time writing and debugging code. We write code to clean, transform, scrape and merge data that we want to analyze. This course will focus on the challenges that arise during this process, and thereby enhance our chances of posing new and challenging questions.

I will present data science methods needed for collecting and analyzing real-world data. In addition to core computational concepts, the course will focus on **generating new data** (collecting, scraping, working with [APIs](http://en.wikipedia.org/wiki/Application_programming_interface)), **data manipulation tools** (munging, transforming, cleaning), **visualization tools** (visualizing raw data and model results), **reproducability tools** (git, github), as well as provide an introduction to statistical techniques for predicting and classification.

**Learning Outcome:**  
Topics:

1. Introduction to Programming
2. Introduction to R
3. Reproducible Research
4. Data Manipulation
5. Static Visualizations
6. Data Gathering
7. Statistical Learning

**Literatures:**  
TBD

**Teaching and learning methods:**  
The course will consist of two hours of lectures and one hour of exercises and problem solving per week. The lectures will focus on broad introductions to the topics covered in this course. One hour of exercises a week is not a large amount of time for learning how to code. We will use some of this time like development meetings: going over assignments, having detailed code reviews of various forms, and discussing blocking issues and potential solutions.

**Academic qualifications:**  
There are no hard requirements, but students are expected to have an interest in some subset of: statistics, econometrics, linear algebra, and a scripting language (we will use R in this course)

### Exam

**Exam registration requirements:**  
Students are expected to complete at least 2 out of 3 mandatory assignments.

**Type of exam:**  
For the exam students are expected to pose an interesting social science question and attempt to answer it using standard academic practices including original data collection and statistical analysis. The project could be a vital preparation for your BA thesis. The project has three presentation outputs designed to present your research to multiple audiences. The first is a oral presentation (10 minute maximum). The second is a standard academic paper (6,000 words maximum) that is fully reproducible. The third is a website designed to present key aspects of your research in an engaging way to a general audience.

**Aid:**  
This course is focused on introducing students to a modern data science workflow. Therefore all aid is allowed.

**Marking scale:**  
The 7-point grading scale.

**Censorship:**

**Re-exam:** Will be as the usual exam.

**Workload in the period:**

* Amount of lectures: 2 hours a week
* Amount of exercise: 1 hour a week
* Amount of preparation to classes: ?
* Amount of hours to write project assignment: ?