There are 2 choices for the final project – every student is expected to choose one and to base the developed analysis plan on.

**Option 1 (cross-sectional data):**

Based on the NHANES dataset that can be downloaded using this code: [Loading of NHANES dataset.R,](https://yu.instructure.com/courses/39395/files/1228437/download?wrap=1)students are asked to develop a hypothesis and use simple analytic procedures to test this hypothesis. Please also read the accompanying documentation of the dataset carefully ( [NHANES.pdfPreview the document](https://yu.instructure.com/courses/39395/files/1228438/download?wrap=1)) - please disregard weight and sampling strata for this assignment.

The hypothesis shall include the analytic testing of a physiological relationship the student is familiar with. The National Library of Medicine Pubmed.gov ([https://www.ncbi.nlm.nih.gov/pubmed/ (Links to an external site.)](https://www.ncbi.nlm.nih.gov/pubmed/)) can be used to identify hypotheses that have been tested in the NHANES data and used for inspiration for developing own hypotheses. A glossary of the included parameters is provided on the following page.

**Option 2 (longitudinal, experimental data):**

The provided simultated data ([data simulated RCT 2020-02-01.csv) Preview the document](https://yu.instructure.com/courses/39395/files/1202139/download?wrap=1)contains simulated (no real patients) data from a virtual randomized controlled trial. The diverse variables describe a before (i.e. baseline) and after period (i.e. follow up), an “intervention” parameter describes intervention versus control group. Students are asked to “invent” the trial with an established (patho-)physiological association, develop a study hypothesis for the conducted randomized controlled trial and to compare a) before and after period and b) control versus intervention arm using the methods taught in the course.

The students are asked to approach either project with creativity and an understanding where to find information (i.e. internet), how to approach and manage a dataset and how to formulate aims, null-hypotheses and how to approach data analytically using R. An analysis plan with a short background should be developed (sample on the following pages) and the final class will serve as the forum to present hypothesis and results in a brief presentation (<10 minutes) with a Q&A with the entire class.

Parameters in the dataset: age, gender, black race, Hispanic ethnicity, creatinine, systolic and diastolic blood pressure, weight, height

**Final Report (outline example):**

Students are expected to structure it in the format of a scientific report with references in format of APA 6th Edition. Structure as follows:

Background

Methods

Results

Discussion & Conclusion

References

**Analysis plan (outline example):**

Background and Rationale/Unmet Need

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Study Aims

Primary Study Aim

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Secondary Study Aims

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Study Hypotheses

Primary Outcome

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Secondary Outcome

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Data Extraction and Analysis Plan

Study Cohort Definitions

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Analyses to perform

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Example:  Descriptive statistics of baseline parameters:

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Example:   Statistical Analysis