\*\*\* Data task for CBS RA candidates (February 2019)

\* Overview

Summary/setting of dataset

We asked around 200 students a series of questions relating to their perception of inequality in income and wealth, their self-perception in the income and wealth distribution, and their redistributive preferences. The answers are attached in raw format. The goal of this exercise is to process the raw data to analyze the

Hypothesis: relation between inequality perceptions, self-perceptions, and redistributive preferences.

\* Detailed task description

(1) Some of the free-text answers need cleaning, both in terms of format and also in terms of contents. For example, there might be some crazy outliers that we need to drop or winsorize.

(2) Summarize the data using histograms of the responses to each individual question. Each graph should contain key information such as number of observations in each response category (when discrete) or the density (when continuous), etc.

(3) Show cross-tabulations or conditional histograms that show the distribution of responses to one question conditional on the response to another question taking on some value. The key cross-tabulation is that between redistributive preferences and perceptions of inequality.

(4) The relation between redistributive preferences should be investigated using a multivariate regression analysis by means of ordinary least squares. All results should be presented in tables with multiple columns or panels, as is common in many professional Economics journals (e.g., the American Economic Review).

(5) Write up your key findings in no more than 2 paragraphs, around half a page.

(6) Your creativity is highly valued. Feel free to suggest and implement additional analyses.

[\* Important notes

(1) All analysis must be done in a statistical program (e.g., Stata, R, Python, etc.). Excel is not permitted.

(2) The final report must be written up in PDF format. A TeX-compiled document (e.g., LyX, TeXShop, etc.) is preferred but not required.]

(3) Graphs and tables must be formatted to make sense. Make sure that response values are sorted in a logical manner and labeled appropriately.

(4) The final document, including graphs and tables, should also be neatly formatted in such a way that it could be circulated to professors and researchers at other universities.

\* General comment

We deliberately give minimal instructions as to what exact graphs and regressions to pursue. A big part of your task as a research assistant (RA) and future researcher is to be an independent thinker and to figure out these things as much as possible. For the purpose of this task, we do not allow for any communication between the RA and the professors, but of course for future RAs on the job this will be encouraged.