README for Replication Package for "Evolving Returns to Personality"

Section 1: Accessing the Data

Our study uses datasets from Statistics Finland, which are not included due to licensing restrictions. In 2024, researchers can access the data via a remote desktop connected to Statistics Finland's virtual databases. To access the data, follow these steps:

1. **Obtain Data Access Permission from Statistics Finland:**
    - Contact Statistics Finland's Research Services (https://stat.fi/index_en).
    - Apply for permission to access these datasets (typically requires affiliation with a Finnish research institute or university):
        - **FOLK_perus**: Basic population data (e.g., earnings). Files used:
            - folk_20112018_tua_perus19tot_3.dta
            - folk_20012010_tua_perus19tot_3.dta
            - folk_19872000_tua_perus19tot_3.dta
        - **FOLK_jaksot työsuhde**: Employment relationships, firm identifiers, industries. Files used:
            - tyosuhde_1.dta
            - tyosuhde_2.dta
    - **EDUC_YTL**: High school final exam results (high school test scores)
    - **EDUC_TYHR**: Upper secondary education joint application system register (9th grade GPA)
    - Crosswalk to map Finnish Defence Forces (FDF) person IDs to Statistics Finland's current person IDs (Tk_tutnro_crosswalk.dta in data environment A72).
2. **Obtain Data Access Permission from the FDF**
    - **FDF psychological data**: Apply for permission from the Finnish Defence Forces' Defence Command. Begin analysis with persons.dta file created for:
        - Jokela, M., Pekkarinen, T., Sarvimäki, M., Terviö, M., and Uusitalo, R. (2017). "Secular rise in economically valuable personality traits." Proceedings of the National Academy of Sciences, 114(25), 6527-6532.

Section 2: Contents of *data_input*, *figures* and *tables* Directories

Before proceeding with the replication process, note that the following files are already present in the *data_input* directory. These files are included in the journal submission because they are not sensitive:

1. **cpi.csv**: Consumer price index data for inflation adjustment.

2. **LUOKITUSAVAIN_AML1980_AML2001.dta**: Classification key for occupational codes.
3. **xwalk_county_subregion.dta**: Crosswalk for county to subregion mappings.
4. **xwalk_industry_3digit.csv**: Crosswalk for 3-digit industry codes.
5. **xwalk_occ_0110.txt**: Crosswalk for occupational codes from 2001 to 2010.
6. **xwalk_occ_1001.txt**: Crosswalk for occupational codes from 2010 to 2001.
7. **xwalk_occupations_2010_to_manual.csv**: Crosswalk for 2010 occupational codes to manually harmonized codes.

For your convenience, the target files to be replicated are already in the *figures* and *tables* folders.

Section 3: Replicating the Results

To replicate the results of our study, follow these steps:

1. **Setup Your Environment:**
    - Ensure you have R and Stata installed on your system.
    - Install the necessary R packages indicated in the code.
    - Ensure the file paths to the original data files described in Section 1 are correct in your remote desktop.
    - Ensure the file paths to *data_input*, *data_prep*, and *data_output* are correct in your remote desktop.
2. **Run the Harmonization and Preparation Scripts:**
    - Prepare the data by running the following Stata do-files in sequence. Output from these steps are saved in the *data_prep* folder:
        - a00 - build_xwalk_prepare.do
        - a01 - build_persons.do
        - a02 - build_person_year_prepare.do
        - a04 - build_person_year.do
3. **Data Tidy and Sample Selection.** Data from these steps are saved in the *data_output* folder:
    - Run 00 - harmonize_xwalks.R to harmonize crosswalks.
    - Run 01 - tidy_persons.R to tidy the individual data.
    - Run 02 - tidy_person_years.R to tidy the person-year data.
    - Select the sample by executing 03 - sample_selection.R.
4. **Analysis and Figures.** Data from these steps are saved in the *figures* folder:

- Generate figures by running 04 - analysis_figures.R.
- Conduct quantile analysis with 05 - analysis_quantiles.R.
- Execute 06 - bundles_revision.R to analyze the earnings growth in different personality trait bundles.

5. **Tables and Descriptives:**
    - Create tables for the manuscript using 07 - analysis_tables.R.
    - Generate descriptive figures with 08 - descriptives.R.
6. **Robustness Checks:**
    - Perform robustness checks using 09 - robustness.R.

By following these steps, you will be able to replicate all analyses and figures reported in our paper, "Evolving Returns to Personality".