GUIDE FOR RESEARCHERS USING BANCO DE PORTUGAL

MICRODATA RESEARCH LABORATORY (BPLIM) DATA

This version: April 20, 2021

1   General Remarks

The Banco de Portugal Microdata Research Laboratory (BPLIM) is part of the Economics and Research Department of Banco de Portugal (BdP) and was created with the objective of facilitating to researchers the access and use of micro datasets about the Portuguese economy. Primarily, these are datasets owned by BdP, but there are other datasets supplied by third parties. Data can only be accessed on BPLIM's dedicated servers. BPLIM has two servers. The **internal server** is primarily intended for use by BdP own researchers (internal researchers) while the **external server** has remote access capabilities and is intended for external researchers. Internal researchers working on the internal server are not bound by any restrictions in terms of data access but restrictions apply to external researchers. Below we detail all relevant aspects involving data access at BPLIM.

2   Data Available for Researchers

2.1   What are the characteristics of BPLIM's datasets?

All BPLIM datasets are stripped of elements that allow for direct identification of companies, banks or individuals. Whenever possible, the datasets contain unique unit identifiers common across datasets: examples of these are *tina* – the tax identification number anonymized for companies – and *bina* – the bank identification number anonymized. Datasets are available in Stata format, are stored in an efficient way that minimizes file size and follow BPLIM's naming convention. Labels are applied to all variables and value labels to all categorical variables. Whenever possible labels can be displayed in Portuguese and English and datasets are accompanied by a manual and a set of auxiliary files that succinctly describe the data. BPLIM datasets may also have companion script files which calculate additional variables or harmonized variables that guarantee comparability over time and across datasets. Datasets are updated regularly based on a data extraction ("data freeze") at a specific point in and a versioning system is applied to reflect any changes to the data set.

2.2   What data are available?

The complete list of datasets including a short description of each one is available in Appendix A. In BPLIM's website you can find a list of the latest version of each dataset along with all updated information including manuals and supporting documents.[1] More detailed information for each dataset (codebook and descriptive statistics) can only be obtained once you are given access to the server. See the External/Internal server manual.

---

[1] The source code for the Data Manuals is available on Github.

3   Data Access

3.1   Who can gain access to the data?

Access is only allowed to researchers who intend to use the data for scientific purposes. Internal users (those with a contractual link to the BdP) can be given access to all datasets available at BPLIM. Access by external researchers is subject to several restrictions. Appendix A summarizes the access restrictions imposed to each dataset.

3.2   How can researchers request access to the data?

Access to BPLIM data implies the submission of a project. The proposal must: (1) contain a short description of the research project; (2) identify all researchers involved in the project and their affiliations; and (3) identify the datasets, timeframe, and variables required. All external researchers with access to the data are required to sign a confidentiality agreement and send a copy of their vitas. A copy of the required documents can be found in BPLIM's website. If the project consists of a master or doctoral dissertation then the supervisor(s) has to be identified and must also sign the confidentiality agreement. BPLIM staff will work with the researcher(s) to identify the required datasets and, if needed, to build a customized dataset.

3.3   Approval process

Upon submission of all documentation the project will be evaluated to ensure that it addresses a legitimate research question. Compliance with the guidelines is also an important factor for recurring external researchers (those that have already participated in BPLIM projects). Once the project is approved data access will be granted and summary information about the project will be posted on BPLIM's website.

3.4   How can the data be accessed?

When applying for the project the researcher must specify if he/she wants to access an account in the **internal** or the **external server**. Accounts open at the **internal server** can only be accessed at the installations of BdP (on-site access) either at Lisbon or Porto. BdP internal researchers can login into BPLIM's internal server from their desktop using their network login credentials (this requires installation of specific software[2]). Users not affiliated to BdP will be considered external researchers. They will be assigned a login and password for BPLIM's internal server and will be granted access to a terminal from which it is not technically possible to transfer, download, copy, paste, or print any data. There is a limited number of terminals available so external researchers must book their visits well in advance.

If the account is on the **external server** then the data can be accessed remotely using a safe connection (we use the _nomachine_ software). With this connection it is not possible to exchange files between the external server and the local computer (similar to the terminal access to the internal server). In special circumstances explained below the

---

[2] BdP researchers should contact the Information Systems Department (DSI) through the 20500 extension number and request installation of the Linux emulation software.

external researcher may be granted access to the data without having an account at the server (**surrogate access**). In this case the external researcher will not have any direct contact with the data and BPLIM staff will act as a proxy for the researcher regarding data access. This means that BPLIM staff will execute the scripts written by the external researcher and share the outputs with the researcher after validation.

3.5   Which are the restrictions imposed on access to BPLIM's datasets?

Internal researchers are able to access all datasets available at BPLIM provided the access is through the internal server but they will have the same access conditions as external researchers if accessing data through the external server.

The access policy for an external researcher is more restrictive. External researchers are granted access to all datasets that can be placed in the external server (these are the datasets classified as "remote" in the column designated "external research access" for the tables in Appendix A). However, if working in a joint project with internal researchers they may access additional datasets. These datasets are identified with the "co-authorship" classification and necessarily have to be placed in the internal server.

Datasets with the classification of "no" in the column "external research access" cannot be accessed by external researchers under any circumstance – data access in joint projects is by internal researchers only. In very particular situations, in projects that do not have internal researchers, it may be possible for an external researcher to gain surrogate access to a dataset. These datasets have the "surrogate" classification (see subsection 3.7).

Datasets made available to external researchers are subject to a confidentiality classification as follows: **low**, **medium**, or **high**. If the variables in a dataset are classified as **low** then the data is only anonymized. However, if a dataset is classified as **medium** then the dataset will contain variables with "modified" values, not the original values. Datasets classified with a **high** level of confidentiality contain variables with "modified" values and are restricted to a subsample of the observations. However, the external researcher can always request to have his/her scripts run on the original datasets.

3.6   How are projects that involve co-authorship between BdP internal researchers and external researchers handled?

In joint projects where the external researcher does not have access to the original data the internal researcher has the responsibility to ensure that the information shared with his/her external co-authors respects the confidentiality requirements associated with the data.

Projects where BdP internal researchers and external researchers have access to the data are designated "mixed projects". If the mixed project is implemented in the external server then the distinction is irrelevant and all researchers are treated as external researchers. However, if the project is in the internal server BPLIM will open an account for the BdP internal researcher(s) and place there all the data requested. External co-authors can be added

as users if the datasets made available in that account have a confidentiality classification of "low". If the account contains datasets with a higher classification level, then the internal user may request the creation of a second "parallel" account. This second account will have both the internal and external researchers as users but the data available will respect the requirements imposed on access for external researchers. It will be the responsibility of the internal researcher to run all scripts on the original data placed on the first account that only he/she has access to. It will also be his/her responsibility to make sure that external researchers do not have any type of access to confidential data; in particular, external co-authors cannot access the project account with the original data, the desktop computer of the internal co-author, or any logs that might contain confidential data. In a mixed project all interaction with BPLIM has to be done via the BdP internal researcher.

3.7 How can an external researcher gain surrogate access to a dataset?

External researchers not involved in joint projects with BdP internal researchers can still gain access to datasets considered to be more sensitive. Appendix A identifies the datasets that can be accessed this way. To gain surrogate access the external researcher needs to submit a detailed project to BPLIM. The project will be evaluated according to the relevance of the topic for the research agenda of BdP. Only projects deemed relevant will be granted surrogate access. If BPLIM decides to support the project then the external researcher will be assigned a data expert at the microdata laboratory (BPLIM) who will interact with him/her and will help prepare and run the scripts on the original data. However, coding is the ultimate responsibility of the researcher and BPLIM will not validate or certify the scripts written by the researcher. External researchers should work in close cooperation with BPLIM staff to ensure that they are obtaining the envisioned results and are encouraged to discuss their findings with BPLIM staff. It is highly recommended that external researchers initiate their project with a short-stay visit in BPLIM. During that time they are expected to discuss their research with BPLIM staff and to gain a good understanding of the complexities of the data. Other visits, during the course of the project, are also encouraged. All outputs shared with the external researcher are subject to the usual disclosure restrictions. Access conditions for external researchers are not waived in any circumstance.

3.8 How can external researchers work with modified data?

To preserve confidentiality of the data BPLIM will often supply modified versions of the original datasets to external researchers. These datasets are made available to facilitate the elaboration of the scripts that manipulate/analyze the data. However, results of analysis performed on datasets with modified variables are not valid for research purposes. Once external researchers finalize their code they can always request that the scripts be run on the original data.

BPLIM can use three different strategies to create modified datasets: perturbed, shuffled, or randomized. A **perturbed dataset** is a dataset that contains "perturbed" variables which are variables where the original values have been replaced by different values which are correlated with the original values. A perturbed file is easily identified because the file name will contain "_P_". A perturbed variable is also clearly identified because its name will always end in "_p". A **shuffled dataset** contains "shuffled" variables, that is, variables that keep their original values but have been

sorted randomly and independently of the rest of the dataset to break the link for the observations. Similar to perturbed variables the file name will contain "_S_" and the variable names will end in "_s". Finally, a **randomized dataset** is a dataset that is both perturbed and shuffled. Randomized variable names end in "_r" and the file name contains "_R_".[3]

If the project involves a co-authorship with internal researchers at the bank ("mixed project") then the responsibility of running the scripts on the original data falls upon the internal researcher since he/she has access to the original data. If the project does not involve internal researchers then BPLIM staff will run the scripts on the original data. This is a process that takes time as we first need to replicate the results with the modified data before moving on to the original data. To speed up the process researchers working with modified data are encouraged to follow the basic rules outlined in Appendix B. In case of any doubts regarding this process, external researchers should contact BPLIM staff directly before writing any script.

3.9   Transfer of external files

BdP internal researchers working in the internal server can freely copy files to and from their account in the internal server and thus are free to merge BPLIM datasets with other data. BPLIM will assist with the anonymization of external datasets to ensure compatibility among datasets. Note that BPLIM identifiers (eg: *tina* and *bina*) are specific to a project and may not be valid if files are exchanged between accounts. If the account is shared with external researchers ("mixed project") then the internal researcher has to respect the guidelines that apply to external researchers. At the request of the internal researcher BPLIM can anonymize and/or modify the external datasets intended to be used in mixed projects.

External researchers may also request that external files are placed on their accounts. The origin of these files must be clearly explained and a justification for their use must be provided. The researcher should also make sure these files can be legitimately used for that purpose. BPLIM staff will assist with the merging of the external datasets with BPLIM datasets. External datasets typically contain aggregated data but it may be possible to add external datasets with finer granularity. In this latter case additional measures may have to be undertaken to ensure that the confidentiality of the data is preserved. These situations will be analyzed case by case and discussed with BPLIM staff.

---

[3] Appendix B provides advices for researchers working with modified datasets.

4 Statistical Software

4.1 Statistical Software

Once the account for the project is created the researcher(s) will have access to a Linux environment that has Stata, *R*, and *Julia* available. Templates for writing code are also placed in the account and researchers should, as much as possible, adhere to the conventions described on the templates. A number of packages that extend the capabilities of the base distributions have been pre-installed. BPLIM staff will assist if the researcher needs any additional package. Installed packages will be placed in the "tools" folder of the project and are available only for that specific project.

4.2 Additional Tools

BPLIM staff has created several tools (Stata packages) to facilitate the tasks of the researcher. Some of these tools are specific to BPLIM datasets while others are of general utility. All of these tools are available on BPLIM's servers. Because we want to be transparent about the coding and versioning we make all tools also available on Github. Users are welcome to suggest improvements or add their own contributions. Tools which are of generic use can be installed directly from Github on any internet connected machine.

5 Output Extraction

5.1 Can the researcher transfer any files from the server?

External researchers are never allowed to transfer files to/from BPLIM's servers. The same policy applies to BdP internal researchers working on the external server.

However, BdP internal researchers are free to transfer files to and from their accounts in the BPLIM internal server. When an account is created for an internal researcher in the internal server the project area from the internal server is mapped to the researcher's personal desktop. This means that internal researchers can easily transfer files to and from BPLIM's internal server including data and output logs. In case the internal researcher is working with external co-authors it will be his/her responsibility to ensure that all files shared with the external co-authors respect the guidelines imposed for the release of information.

5.2 What restrictions apply to the release of output logs?

As a general principle BPLIM will never check for the "correctness" of the scripts that researchers use to produce the logs and declines any responsibility for any errors or imprecisions that are present in researcher's code. This is the sole responsibility of the researcher. Output logs should never contain information that reveals individual records of the dataset, and thus should only contain aggregate-level information. This means that listings of individual records, tables with cells with values obtained from manipulation of three or less observations, statistical measures with standard errors of zero, minimum and maximum values, etc., are not allowed. Plain text files (including Latex and comma or tab

separated values) are the preferable format but other formats may be accepted provided that data content can be easily checked (e.g. small Excel files). Graphical outputs should be produced in the ".png" format. In mixed projects the responsibility of checking adherence to these principles falls upon the internal researcher. If requested by the internal researcher, BPLIM will assist with that process. In projects involving only external researchers, it will be the responsibility of BPLIM staff to verify conformability of all outputs. Output verification depends on staff availability and may take longer in periods when the workload is higher. BPLIM staff will only answer requests for output extraction sent by email and will take as long as needed to ensure that all confidentiality requirements are safeguarded. Researchers should keep their output requests to a minimum and, as much as possible, these should be of final outputs. In the case where researchers are working with modified datasets (eg: perturbed files) and want to look at intermediary results obtained from original datasets it may be possible to place the logs in a folder in the researcher's account instead of emailing the results.

5.3   What happens if the researcher violates the rules?

BPLIM assumes that all researchers are operating in good will and they should strive to follow the dispositions set forth on the signed "Declaration on Confidentiality and Use of Data". If a researcher behaves in a way that is considered inappropriate BPLIM reserves the right to cancel all projects in which the researcher is involved.

6   Replicability and Data Archiving

6.1   Replicability of work

BPLIM will work with third-parties such as data editors, certification services, or individual researchers to provide conditions for replication of results for individual projects. The replication process consists of opening an account for the "replicator" and providing him with the same BPLIM datasets of the project. Additional files from the project may be provided if there is a request for their archival (see below). To facilitate replication researchers should unequivocally identify all datasets used in the analysis. All other needed files should be provided by the "replicator". BPLIM will work directly with data editors or certification services (e.g. Cascad) to evaluate the best approach to implement their replication protocol. In the case of individual researchers willing to act as "replicators" they will have to go through the standard process of submitting a research project.

6.2   Archiving

By default, once a project is closed BPLIM will keep a copy of all syntax files (e.g. text files with "do", "R", and "jl" extensions) plus all files found in the "*initial_dataset*" and the "*tools*" folder. A copy of the syntax files will be sent to the researcher who should verify that the list is complete. All other files will be deleted unless the researcher explicitly requests that certain external file(s) (see subsection 3.9) are archived. Only in exceptional and well justified circumstances will BPLIM agree to archive intermediate data files created by the researchers. It is the responsibility of

the researcher to ensure that all files needed for proper replication of the results are archived with the project. Archives are kept for ten years since the closing of the project.

7 Publication of Research Papers

7.1 Citations and research outputs

All BPLIM datasets should be cited according to the information provided in the data manual. If available, the Digital Object Identifier (DOI) should be referenced. When the topic analyzed by the external researcher(s) bears special relationship to BdP's statutory tasks the researcher is encouraged to discuss their results with BdP staff. In this case, BdP may ask to see a copy of the work prior to any public release and may provide suggestions regarding the research project. Moreover, in situations where data access is granted based on the relevance of the topic ("surrogate access") researchers are not only encouraged to discuss their results with BdP staff but must also seek BdP approval prior to any public release of their results. As soon as available, researchers are required to send to BPLIM a copy of all research outputs (working paper, conference proceedings, paper, thesis, etc.) related to the project.

8 Contacting BPLIM

8.1 How can I contact BPLIM?

The preferred way to contact BPLIM is through email: BPLIM@bportugal.pt. For projects already ongoing the subject line should always include the project reference (eg: p###_Surname). If necessary you can contact us at:

Address: Banco de Portugal

Microdata Research Laboratory

Rua do Almada, 71

4050-036 Porto

Portugal

APPENDIX A

Datasets Available at BPLIM

1. Central Balance Sheet Database

The Central Balance Sheet Database is constructed and made available by Banco de Portugal and provides economic and financial information on non-financial corporations operating in Portugal. It contains annual data from 2006 onwards and is mostly based on information reported through *Informação Empresarial Simplificada* (IES, Simplified Corporate Information). Until 2009 the data were reported according to the *Plano Oficial de Contabilidade* (POC, Official Chart of Accounts). In 2010, a new accounting system, the *Sistema de Normalização Contabilística* (SNC, Accounting Standards System) was put in place. The main goal of SNC is to make the reporting of accounting information compliant with International Accounting Standards (IAS). Thus, all data reported after 2009 follow the SNC. This is a structural change that is reflected in the way the dataset is organized and poses some challenges in the comparability of the variables over time. BPLIM provides two different sets of files: the annual firm-level set and the firm harmonized panel. The first data set contains the data as reported while the harmonized panel contains variables that were made consistent over time. For more information see the data Manuals.

Name	Data Range	Confidentiality Classification	External Researcher Access	DOI
Firm Level:				
Yearly Data:				
Apr2017 Extraction	2006-2015	Low[1]	Remote	-
Jun2018 Extraction	2006-2016	Low[1]	Remote	-
Jun2019 Extraction	2006-2017	Low[1]	Remote	https://doi.org/10.17900/CB.CBA.Jun2019.V1
Jun2020 Extraction	2006-2018	Low[1]	Remote	https://doi.org/10.17900/CB.CBA.Jun2020.V1
Harmonized Panel Data:				
Apr2017 Extraction	2006-2015	Low	Remote	-
Jun2018 Extraction	2006-2016	Low	Remote	-
Jun2019 Extraction	2006-2017	Low	Remote	https://doi.org/10.17900/CB.CBHP.Jun2019.V1
Jun2020 Extraction	2006-2018	Low	Remote	https://doi.org/10.17900/CB.CBHP.Jun2020.V1

(1) There are two variables that if requested by the external researcher will be made available in the perturbed format.

2. Central Credit Responsibility Database

The Central Credit Responsibility (CCR) database reports credit supply by all credit-granting institutions in Portugal. Data is collected monthly with the objective of supporting participants in the risk assessment of credit concession. Banco de Portugal collects information at the credit exposure level but produces several datasets with different levels of granularity: exposure, firm-bank-level and firm-level data. The data sets can be aggregated to quarterly or annual values. For the firm-bank level data BPLIM produces a harmonized panel which can be aggregated to the firm level. All datasets are updated annually. For more information see the data Manuals.

Name	Data Range	Confidentiality Classification	External Researcher Access	DOI
Firm Level:				
Monthly/Yearly Data:				
Apr2017 Extraction	Jan. 1999-Dec. 2016	Medium	Remote	-
Apr2018 Extraction	Jan. 1999-Dec. 2017	Medium	Remote	-
Apr2019 Extraction	Jan. 1999-Aug. 2018	Medium	Remote	-
Panel Data Harmonized:				
Apr2018 Extraction	Jan. 1999-Dec. 2017	Medium	Remote	-
Apr2019 Extraction	Jan. 1999-Aug. 2018	Medium	Remote	https://doi.org/10.17900/CRC.FRM.Jun2019.V1
Firm-Bank Level				
Monthly/Yearly Data				
Apr2017 Extraction	Jan. 1999-Dec. 2016	Medium	Co-authorship / Surrogate	-
Apr2018 Extraction	Jan. 1999-Dec. 2017	Medium	Co-authorship / Surrogate	-
Apr2019 Extraction	Jan. 1999-Aug. 2018	Medium	Co-authorship / Surrogate	-
Panel Data Harmonized:				
Apr2018 Extraction	Jan. 1999-Dec. 2017	Medium	Co-authorship / Surrogate	-
Apr2019 Extraction	Jan. 1999-Aug. 2018	Medium	Co-authorship / Surrogate	https://doi.org/10.17900/CRC.FRMBNK.Jun2019.V1
Exposure Level:				
Monthly/Yearly Data				
Apr2017 Extraction	Jan. 1999-Dec. 2016	Medium	Co-authorship / Surrogate	-
Apr2018 Extraction	Jan. 1999-Dec. 2017	Medium	Co-authorship / Surrogate	-
Apr2019 Extraction	Jan. 1999-Aug. 2018	Medium	Co-authorship / Surrogate	https://doi.org/10.17900/CRC.EXP.Jun2019.V1

3. Bank Balance Sheet

The Monetary Financial Institutions (MFIs) Balance Sheet Database reports detailed information on the assets and liabilities of the MFIs operating in Portugal. This data set can be of interest for both research and the monitoring and evaluation of emerging risks in the Portuguese financial system. The dataset is updated annually. For more information see the data Manual.

Name	Data Range	Confidentiality Classification	External Researcher Access	DOI
Monthly Data:				
Apr2017 Extraction	1997-2016	Medium	Remote	-
Jun2018 Extraction	1997-2017	Medium	Remote	-
Jun2019 Extraction	1997-2018	Medium	Remote	-
Jun2020 Extraction	1997-2019	Medium	Remote	https://doi.org/10.17900/BBS.Sep2019.V1

4. Quadros de Pessoal

Quadros de Pessoal (QP) is available through a Protocol celebrated between Banco de Portugal and the Ministry of Labour and Social Solidarity and may only be used in projects involving internal researchers of BdP. It covers all public and private firms with at least one employee. QP contains yearly data on firms, establishments, and their workers with detailed information on worker characteristics. For more information see the data Manual.

Name	Data Range	Confidentiality Classification	External Researcher Access
Firm Level			
Yearly Data	1982-2013	Low	No
Establishment Level			
Yearly Data	1982-2013	Low	No
Worker Level:			
Yearly Data	1982-2013	Low	No

5. Fast and Exceptional Enterprise Survey – COVID-19

The Fast and Exceptional Enterprise Survey – COVID-19 (COVID-IREE) was launched by Statistics Portugal (INE) and the Banco de Portugal aiming to identify the main effects of the COVID-19 pandemic on key aspects of the enterprises activity, such as firm's turnover, workforce, prices, credit conditions and the use of Government support measures. This survey was addressed to a representative sample of non-financial firms located in Portugal. The data started being collected with a weekly frequency on April 6-10, 2020 and refers to a fortnight from May 2020 onwards. Although the survey was suspended after July 2020, two new editions were implemented, in November 2020 and in the first fortnight of February 2021 given the evolution of the pandemic situation. For more information see the data Manual.

Name	Data Range	Confidentiality Classification	External Researcher Access
Firm Level:			
Weekly/Fortnightly/Monthly Data:			
Jul20 Extraction	Edition 15 (6-10 April 2020) to Edition 23 (1st fortnight of July 2020)	Low	Remote
Nov20 Extraction	Edition 24 (November 2020)	Low	Remote
Apr21 Extraction	Edition 01 (1st fortnight of February 2021)	Low	Remote

6. Historical Series of Portuguese Banking Sector (SLB)

The Historical Series of the Portuguese Banking Sector Database (SLB) reports, on a consolidated basis, a wide range of series on bank's financial statements (i.e., balance sheet, income statement, and solvency), loans to customers, interest rates, human resources, branch network, and payment systems. The data is collected and assembled by a working group at Banco de Portugal which was established at the end of 2017 with the objective of constructing historical series on the Portuguese banking sector. The SLB database covers the period from 1990 to 2019 and is available at yearly frequency. For some tables (i.e., balance sheet, income statement, solvency, loans to customers, and interest rates), data is also available at quarterly frequency. The dataset is updated annually and the coverage of financial institutions varies depending on the data table. For more information see the data Manual.

Name	Data Range	Confidentiality Classification	External Researcher Access
Banking Group/Bank Level:			
Yearly/Quarterly:			
Oct20 Extraction	1990-2019	Medium	Remote

APPENDIX B

Replication on Modified Datasets

External researchers working with modified datasets may ask at any time that BPLIM staff runs their scripts on the original data. However, to expedite this process external researchers should adhere to a set of rules. Below, we describe these rules (assuming that the code is written in Stata and therefore that the researcher is preparing a "do" file, but the same logic applies to other statistical languages).

1. The scripts should be based on the template provided. Importantly, the path for all folders must be defined right at the beginning of the "do" file (even if you call other "do" files). See the example in the template.

2. Whenever using modified variables you must always code based on the modified datasets provided by BPLIM that were placed in the *initial_dataset* folder. If you calculate intermediary datasets based on the modified data they will have to be recreated each time you rerun your code.

3. Your code should be able to run in a single execution. This means that you can either have a single "do" file or a main "do" file that calls all other "do" files.

4. You must create a log file in text format that proves that your script runs without interruptions from "top to bottom". Only after we successfully replicate this log file do we run the script on the original non-modified data.

5. Avoid using "_p", "_s" or "_r" in the name of variables you create in the script.

6. Never run any conditions based on absolute values of modified variables. These values may not be meaningful once we rerun on the original data.

7. Please make sure that your script is completely reproducible before submitting it. For example, if you are working with random numbers always set the seed.

8. Remember that your outputs should never allow for direct or indirect identification of individual values.