

**BIST/STAT 5505.**  
**Applied Statistics I (Fall 2018)**  
**Small Project**

The small project consists of a data analysis report on a dataset of your choosing *or* on a subset of the “Crime Data”. The complete dataset is available online, at <https://catalog.data.gov/dataset/crime-data-from-2010-to-present>

You are free to download the complete dataset, but it’s large and has many missing values. For convenience you can find a smaller subset on the course website. It contains 500,942 crime records from 2010 to 10/2017 that occurred in the City of Los Angeles. The subset of the data contains the following columns:

- |                       |                      |
|-----------------------|----------------------|
| 1. DR.Number          | 9. Victim.Sex        |
| 2. Date.Reported      | 10. Victim.Descent   |
| 3. Date.Occurred      | 11. Premise.Code     |
| 4. Time.Occurred      | 12. Weapon.Used.Code |
| 5. Area.ID            | 13. Status.Code      |
| 6. Reporting.District | 14. Latitude         |
| 7. Crime.Code         | 15. Longitude        |
| 8. Victim.Age         |                      |

The file “CodeDescriptions.txt” contains descriptions of some of the codes.

If you choose to use a dataset other than the Crime Data that is provided on the course website, your dataset *must* contain multiple variables, with at least one quantitative variable and at least one qualitative variable on which all (or the majority) of sampling units are measured.

In the report, you must document a *thorough* exploration of the dataset using the graphical and numerical summary techniques as well as perform statistical inference procedures discussed in this class.

This is a team project. Every team should be made of two-three students. By Sunday November 4, 2018 at 11:59 pm: Identify a team (2-3 students) and inform your instructor by email, copying your partner/s. Just one member of a team should email the instructor. The typed final report (maximum 5 pages including tables, figures, and references) is to be submitted to the instructor via email by 11:59pm, Friday, **December 7, 2018**. The report needs to be in a Microsoft Word (.doc or .docx) or pdf file. Name the report as LastName\_FirstName\_project.docx(pdf), where LastName (one of the group members) and FirstName (same group member) are filled in appropriately. Only one report per group should be submitted.

The report should include the following components:

- Title Page

This includes title of the project, your names, contact addresses, and email addresses.

- Abstract/Summary Page

This includes an abstract or summary of your work.

- Section 1. Introduction and Description of Data

This section should include, at minimum, (i) clear statement(s) about the research question(s) that you will investigate and (ii) which variables are qualitative and quantitative.

- Section 2. Exploratory Data Analysis

This section should include labeled figures, graphs, and tables that are referenced and discussed within the main text (in Section 2 and possibly elsewhere).

- Section 3. Statistical Inference

This section should include statistical inferences such as confidence intervals, hypothesis tests, nonparametric inference procedures, and/or categorical data analysis techniques. Do NOT use tools and/or techniques not covered in this course (e.g., regression).

- Section 4. Discussion of Results and Concluding Remarks

This section should include a summary of your findings.

- Section 5. Appendix (optional)

This section could include computer code, the dataset itself if you choose your own data, or any other important additional materials.

- References (if applicable)

Cite all sources, if any, used in your analyses/report. Citations should be formatted as follows: Authors (Year). Title. *Journal*, **Volume**, page numbers.