

## Work Plan

### Abstract:

The document aims to provide a work plan for our project proposal for Information Environments (INFM 600) at the University of Maryland. The work plan will specify our draft research questions, planned timeline for completing the project, projected effort allocation and target audience for the analysis.

### Introduction:

The Toxics Release Inventory (TRI) is a publicly available database containing information on waste management activities and toxic chemical releases in the United States. The data spans across the time period from 1987 to 2015, and is available at <https://www.epa.gov/toxics-release-inventory-tri-program/tri-basic-data-files-calendar-years-1987-2015>. It includes all industry groups which are producing more than 25,000 pounds of waste and/or handling more than 10,000 pounds of listed toxic chemicals annually.

### Draft Research Questions:

- 1) The industries/companies/POTWs present in the given vicinity, and how they are contributing to the pollution? We can scale up the data from county to state level and identify potential areas under threat from high levels of pollution. Which state ranks the highest in pollution? Which county in a state contributes to highest pollution?
- 2) Studying the data over a decade, we can analyze how the waste management is being done by the industries? Is the amount of waste increasing or decreasing over time? How much waste is being recycled, how much waste is being converted to energy, how much waste is being treated? Are the companies moving towards better or worse waste management?
- 3) Which industry type is causing the most amount of pollution (by volume/weight)? Which chemical compound contributes to the most amount of pollution? Which industry is producing the most carcinogenic compound? If we can find a dataset related to the amount of cancer related cases/deaths over the USA, is there a direct correlation between the them?
- 4) Which companies are more prone to stack (traditional means) emissions versus fugitive (escaped/leaked) emissions?

**Planned Timeline for completion of the project:**

We have set up a tentative timeline and will strictly adhere to it for successful completion of this project.

**Week 3 (September 14)**

- Team formation

**Week 4 (September 21)**

- Dataset choosing (based on interests and probable research questions which can be answered and dataset allocation.

**Week 6 (October 5)**

- Project proposal and work plan.

**Week 7 (October 12)**

- Understanding the research questions and statistically representing it.
- Designating all the dependent and independent variables for each question.
- Also deciding and finalizing on any other dataset that we may require in order to collaborate our findings.

**Week 8 (October 19)**

- Finding out required variables and redundant variables based on mathematical representation of each question.

**Week 9 (October 26)**

- Cleaning of data and formatting it according to R requirements.

**Week 10 (November 2)**

- Data cleaning documentation draft.
- Team progress meeting.

**Week 11 (November 9)**

- Understanding all the mathematical operations involved in each of the research questions.
- Learning the required syntax in R.

**Week 12 (November 16)**

- Implementing code the R code.
- Final Inferences
- Creation of R script draft.

**Week 13 (November 30)**

- Visualization of data and its interpretation.
- R plot draft.

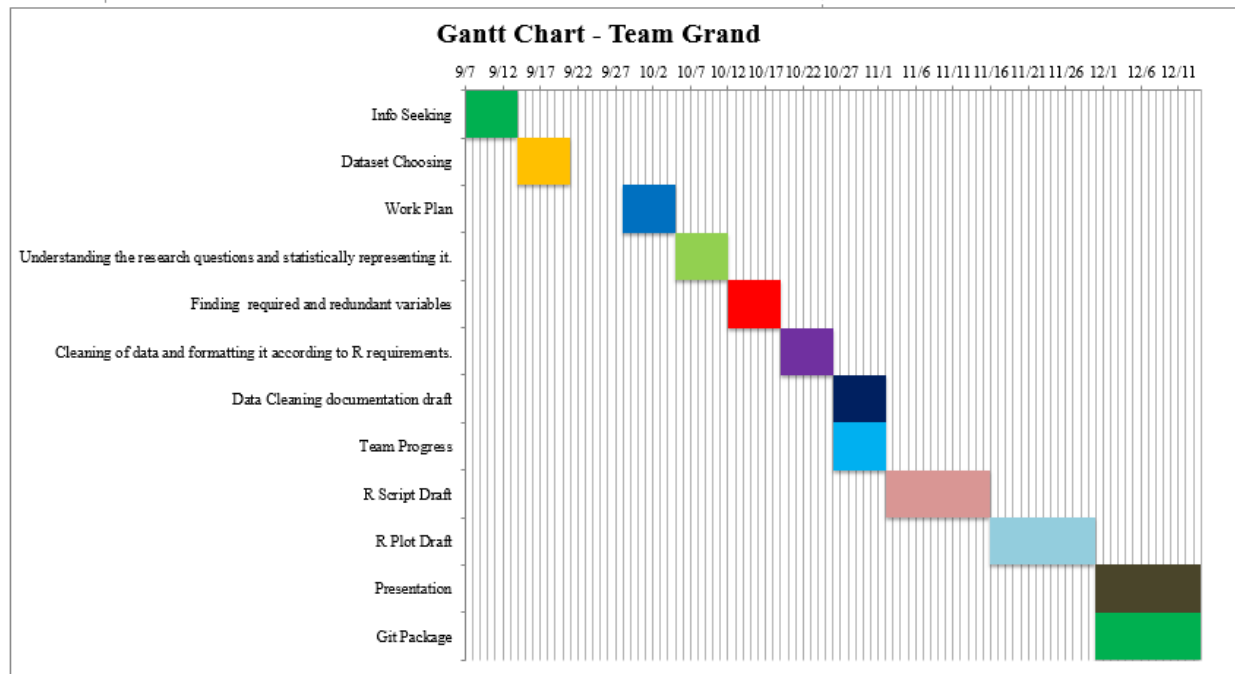
**Week 15 (December 14)**

- Final Presentation
- Git Repository

The same can be represented by a Gantt chart, which is type of bar chart, used to illustrate a schedule. Gantt charts illustrate the start and finish dates of the terminal elements and summary elements of a project. We have generated a Gantt chart in excel which needs a table as data source. The data table and Gantt chart are shown below (also the excel file containing the same is included)

Task Name	Start	End	Duration (days)
Info Seeking	9/7/2016	9/14/2016	7
Dataset Choosing	9/14/2016	9/21/2016	7
Work Plan	9/28/2016	10/5/2016	7
Understanding the research questions and statistically representing it.	10/5/2016	10/12/2016	7
Finding required and redundant variables	10/12/2016	10/19/2016	7
Cleaning of data and form	10/19/2016	10/26/2016	7
Data Cleaning documentation draft	10/26/2016	11/2/2016	7
Team Progress	10/26/2016	11/2/2016	7
R Script Draft	11/2/2016	11/16/2016	14
R Plot Draft	11/16/2016	11/30/2016	14
Presentation	11/30/2016	12/14/2016	14
Git Package	11/30/2016	12/14/2016	14

This table represents the above timeline as mentioned, we use this for developing the schedule using the Gantt chart which can be found below



### Projected Effort Allocation:

All the team members will try to equally divide the work load among them. We would like to utilize this project to maximize our learning and problem solving skills and hence all three of us have decided on being part of every aspect of the project. If one of the team members is having better skills in a given section, others will learn from the team member and try to implement with help. We are planning to schedule weekly virtual meetings wherein we would discuss the previous week's tasks and also chart out the next week's schedule. Since the collaboration will be done on GitHub, we can effectively track everyone's efforts and progress.

### Target Audience:

The target audience which would utilize the dataset and the results obtained from the proposed research questions is:

- 1) The Government of United States
- 2) The United States Environment Protection Agency(EPA)
- 3) The people of the United States
- 4) Oncology Researchers

The government along with EPA, can study the results so obtained, and from the answers, identify areas which are under direct threat from pollution from industries. This can help them set various guidelines which will help in better pollution prevention and better waste management. People of the USA have a right to know under how much threat they are and how much risk they face because of presence of industrial sectors in their vicinity. People can calculate the amount of human risk involved by using the EPAs publicly available Risk-Screening Environmental Indicators model. People of the USA have a right to know under how much threat they are and how much risk they face because of presence of industrial sectors in their vicinity. People can calculate the amount of human risk involved by using the EPAs publicly available Risk-Screening Environmental Indicators model. With the help of this data and research, the Government, EPA and people can make informed decisions.

Word Count: 918