Group Project Diary		
Assigned To	**Tasks**	**Description**
Ahn Joo-Hyun	Topic discussion	- Suggested a topic which are airline delay data set, crash accidents data sets by airport and airlines.
	Data wrangling	- Replaced or deleted some dirty data what we do not use in the future in Julia. - Bound every the-numbers dataset (7 files with movieId after using API for dataset from numbers.com) to do join with IMDB data in R. - Joined data between IMDB dataset (basic, rating, crew) in R. - Created tidy version CSV files to the next step in Julia. - Create new variables from given columns for analysing movie data such as counting words in the title, calculating genre complexity, Rescaling a financial data for easy to read.
	Plotting	- Made graphs for visualisation in Julia. - Analysing the relationship between some factors for example budget, words count, rating and revenue
	Presentation	- Presented visualization part of presentation and conclusion.
	Organisation	- Created a google drive and diary for the group to share ideas and relevant materials. - Draw an overall outline how to proceed to the group project. - Coordinated with a team and conducted regular meetings.
	Prototyping	- Active discussion. Offered ideas about how to combine two different datasets and visualization.
	Report	- Wrote report part related to plots, visualization, conclusion. - Reviewed codes regarding the wrangling and plotting in Julia part. - Summarised R and Julias codes for wrangling and visualisation and made comments for each steps. - Updated project diary. - Consolidated final codes in R and Julia by collecting each parts from group members.
Jia Zhiying	Topic discussion	- Offered two different potential topics, one is to combine weather information with traffic data; another is to scrape and analysis the replies of a popular post.
	Data wrangling	- Optimized the data model from the perspective of tidy dataset and suggest to delete columns with the similar attributes.
	Web scrapping	- Completed web scraping of box office data from boxofficemojo.com, but the amount of data is less, only 790. - The project team decided to scrape more box office data from the-numbers.com. - Worked with Moon to figure out how to get the changed page links.
	Prototyping	- Active discussion, offered ideas about visualization.
	Presentation	- Prepared the presentation and slides about the goals and data sources.
	Report	- Wrote report part related to goals and data sources. - Wrote the project report, consolidation in one file of all parts written by members. - Updated project diary.

Group Project Diary		
Assigned To	**Tasks**	**Description**
Changmin Moon	Topic discussion	- Presented a topic about movie datasets which can lead to prediction of specific upcoming movies. - Regarding the topic, found two main data set sources: imdb.com and the-numbers.com.
	Web scrapping	- Scraped the financial information from the-numbers.com and managed to collect the data about 5786 movies as a single data frame.
	Data wrangling	- Wrangled the movie data collected from the-numbers.com such as omitting unnecessary units. - Wrangled date column into 'yyyy-mm-dd' format - Once stuck in wrangling date column but found the way by using 'ludridate' package.
	Prototyping	- Active discussion. Offered ideas about what unique key should be and what to analyse and present.
	Presentation	- Edited and collected whole PPT material for the presentation. Made the beginning part of presentation. - Presented introduction of the presentation.
	Report	- Summarised R codes for web-scraping, and made comments for each steps. - Wrote report part related to introduction, web-scrapping, and wrangling about the-numbers.com data parts. - Summarised the whole group project into a brief txt file to publish the group project to Github. - Uploaded relevant files of the group project to Github, and shared with other members how to work. - Edited the group project report and added supplementary explanation. And updated project diary.
	Organisation	- Proposed guidelines for presentation and defined progress.
Komlev Viacheslav	Topic discussion	- Offered some topics: Quality of life by cities, Quality of life by countries, energy production and contribution in NZ or worldwide in relation with CO2 emission. - Found data with movies revenues, links saved in common folder on google drive, file DataSetLinks.doc.
	Data wrangling	- Uploaded some tsv files (title_basic, title rating) from IMDB and found the way to read data without errors. - Made some wrangling (filter by year, excluding values which are not used in the project). - Wrote a function which mutate a new column with imdbID. - Created new title_genres table (save data into files title_genres.csv, title_rating.csv).
	Using API	- Wrote an API function which gets movieID from omdbapi.com for financial data from the-numbers.com. - Created several CSV files for further processing by using API function.
	Prototyping	- Active discussion, offered ideas about visualization, API function for joining data.
	Presentation	- Preparation for presentation, data model, wrangling, and API function parts.
	Report	- Added comments regarding the codes for API function and relevant wrangling in R. - Project diary aggregation. - Updated project diary. - Wrote about parts regarding data model, project steps, data wrangling techniques, API function.
	Organisation	- Arising questions about necessity of discussion related to project. (solving problems, topic searching).