# Group Project Proposal

# PieCharts4Life

UCACCMET2J Data Analysis For LAS

17 January 2020

Amber, Daniel, Ronit, Roos, Vivian

## Research Question:

What is the difference in net worth between an actor that acts in drama or action-movies?

Variables: Genre, net worth,
Alternatives/Additions: Number of episodes, number of films, gender

If these genres are not coded specifically, we can change it to what is encoded. The question can be extended by adding on to it with number of episodes/movies this actor played in or their gender. Other things like date of birth, or salary can be included as well. It all depends on how much time we have which ones we want to include. It would also be interesting to consider different kinds of genres and make a ranking of those.

## Timeline:

Throughout the week commits to our GitHub version history will be updated and we will write bits and pieces of the report in Latex. For writing the report, we aim to have most of the introduction and methodology section written alongside or shortly after writing sections of the code. Most of the discussion will follow the coding, in order to draw conclusions based on the acquired results and visualizations.

Link to GitHub: https://github.com/Roos-Visser/Group_Project

| | |
|---|---|
| Monday | Create basic python pipeline -> tidy data |
| Tuesday | Start working in R studio and make python pipeline more complex (if time) |
| Wednesday | Visualisation of the data, make R pipeline more complex |
| Thursday | Prepare presentation, work on Latex report |
| Friday | Presentation, final updates, edits and details of the pipelines and report |

## Work division:

Tasks:
- Creating an outline in Latex
- Exploring the data
- Select appropriate data in PowerShell/Bash
- Filtering and manipulating the data set in Python

- Conversion to csv
- Tidying up the data set in R
- Creating visualizations of the data in R
- Writing report sections in Latex
- Adding details in the Latex report (sources, making it look pretty etc.)
- Making the PPT presentation

**INITIAL THOUGHT PROCESS**

Interesting to look at:

- Gender (how many more women than men… )
- Date of birth
- AlmaMater / University / College / School
- Cause of Death
    - Which is the most common
- Net Worth - good to combine with anything?
    - Or salary
- Ethnicity
- Religion

- NumberOfEpisodes/Films -> net worth
- Genre -> net worth

- Weight of presidents and vice-presidents / prime-ministers / ministers
    - Compare to which party they come from

What is the relationship between murder, salary and net-worth of the victim → are rich people more often murdered than poor people?
What is the relationship between gender, date of birth and salary/networth?

Talk:

Problem: very sparsely encoded

**Research Question:**

**TEAM NAME:**

- **WeStriveToExcel**
- **PieCharts4Life**

This needs to be in there

- short proposal with your ideas and
- how you are going to manage.
- 1-2 pages
- brief introduction to your project so that we can think about the feasibility over the weekend and give you some pointers on Monday.

How to manage:

1) If we are to compare presidents and vice presidents for weight, gender, etc. - there is a need to filter out for the respective positions. For this we have a variable which tells us that. Weight is an integer. Gender a string(?) for which we can select. The next step would then be to which party they relate. Weight could be related to stress level. With more of the former and less time for a 'healthy' way of life (less sportive activity, less qualitatively healthy food).

2) Are rich people murdered more often than 'poor'?
What is the relationship between murder, salary and net-worth of the victim → are rich people more often murdered than poor people? What is the relationship between gender, date of birth and salary/networth?
First, we filter out for actually physical people by negating the "mythology" variable in the data set. Following this there is a need to define what we understand as 'rich' and what as 'poor'; various groupings of income-range are imaginable. Then a division in murdered and not-murdered has to be done, and compared between groups to answer the question.
3)