Data Visualization – Assignment 1

– How do I get rich with movies (or die trying)? –

P. Rüdiger, J. Lukasczyk · H. Leitte – winter term 2017/18

In the first assignment we want you to actively think about visual designs and ways to turn data into graphics to tell a story. We will work on cinema data that you can collect from the web.

On the webpage `http://boxofficemojo.com` you get all related information to movies, like the number of visitors, actors, genres, costs and grosses (grand totals, weekly values, country-based grosses). In this assignment, you will gather the information from this site and visualize it to be able to answer the questions below more easily.

The goal is to make you think about the visual designs you already know and how they can be used effectively. Pay also attention at the many different choices you have in the design process and how they are made available to the user of the software.

We recommend to use Bokeh `https://bokeh.pydata.org/en/latest/`, a Python interactive visualization library. Some sample scripts are given in the material forlder in OLAT. We will keep on working on Bokeh a lot in this course, hence, we recommend to start early. You may also use Microsoft Excel or similar to solve this assignment.

Due date: 08. November 2017, 22:00

Bokeh and computer account: Bokeh is a python library that enables users to quickly design charts in the style of D3.js (without the hazzle of learning javascript and base D3). We have preinstalled all relevant libraries in the computer pool of our group (room 36-223), which is open to students during office hours (Mo-Fri, 9:00 - 17:00), except for times when regular courses take place.

To **request an account**, please enter your RHRK user id into the "Actions → Request login" form in OLAT.

You can also install Bokeh on your own computer/laptop. The homepage gives good support: `https://bokeh.pydata.org/en/latest/docs/user_guide/setup.html`.

Submission instructions:

- Work in **teams of 2**. Register for teams in OLAT under "Actions → Enroll in team" with your partner.
- Please submit your solutions as a **single pdf document** per team ("Submissions → Assignment 1").
- If you are already familiar with python programming, you may look into jupyter notebooks, which have an integration for Bokeh. We will cover this topic later in the course for everyone.
- Present for each exercise your chart. Also add a brief answer to the given questions (one or two sentences are sufficient).

Exercise 1 – Weekly Gross Development — 8 points

In the first exercise, we will analyze the gross of a single movie over time. Go to the Weekly Box Office data (see menu on the left). Select a year you are interested in and sort the table by #1 Movie and pick a movie whose title starts with the first letter of your given name (we want to make sure that people in the course look at a variety of movies).

Starting Bokeh and importing data (3P): Your assignment folder contains the file `filmgross.py`, which gives an initial lineplot of the weekly gross of the movie X-Men: First Class from 2011.

- If you work on a vispool computer, load anaconda once at the beginning of your session:

```
> source /opt/anaconda/anaconda-env.sh
```

- Run the provided python script from commandline (No prove required for submission.):

```
> python filmgross.py
```

- Now update the file to show the data for your movie. Add the resulting image to your submission.

Visualization and analysis (5P): Now improve the baseline visualization to be easier to read and more appealing. Also add the rank of the movie to your chart. Using your visualization, the user shall be able to answer the following questions:

- How is the weekly gross developing over time?
- Is there a correlation with the rank of the movie?
- How long does the movie stay in the Top 10?

Hints: The bokeh documentation gives helpful support:

- Quickstart guide: `https://bokeh.pydata.org/en/latest/docs/user_guide/quickstart.html`
- Tutorials: `http://nbviewer.jupyter.org/github/bokeh/bokeh-notebooks/blob/master/tutorial/00%20-%20Introduction%20and%20Setup.ipynb`

Take care of the annotations/labels of your chart, to enable a fellow student answer the questions above and do related analysis.

Exercise 2 – Market Shares

8 points

Next we will look into the market shares of different studios. Under the menu entry “Indices” → “Studios” you can find the market shares of many studios. Choose a year randomly and visualize the market shares such that you are able to answer the following questions:

- Is there a dominating studio? (4P)
- How does the share compare to the number of movies produced by this studio? (Did they just produce more films or were their films more successful?) (4P)

Exercise 3 – Success of genres

8 points

In the last exercise, we will analyze if a certain genre does particularly well. Pick a genre and visualize when movies in this genre were released und how much gross they made. (Pick at most the top 30 movies).

Questions for the visualization (5P):

- Is there a time, when movies of your genre were very popular?
- Which movie of your genre made the highest gross (when and how much)?
- How likely is it that a film from this genre becomes a great financial success?

Pick the 5 most successful movies from your genre und analyze how they performed in other countries (outside the US). Questions for the visualization (3P):

- In which countries can you make the highest profit with this genre?
- Is the genre famous in a certain region or culture?

Contact: ruediger@rhrk.uni-kl.de, leitte@cs.uni-kl.de (Please only use in urgent cases. Questions are answered after the lecture and during the exercises.)