Welcome to the second problem set of A Primer in Data Science 2017. You will be assigned to a tutorial group on Wuecampus. Your group should upload the solution to this problem set by June 4th 23:59.

Your submission should come in the form of **sufficiently documented** R Markdown files for each question. Please provide a rendered version of the document (HTML / WORD / PDF) as well as the rmd source file. Scoring will be based on correctness, argumentation, presentation and creativity. The question marked with a * is a challenge question. There is not a per se correct solution but rather there are many paths to different great solutions. Surprise me!

If there are any problems feel free to put them up on the Wuecampus forum.

1: Apartments in Würzburg

(a) Getting a list of listings

On wuecampus you find a script *apartmentListings.r*. Load this script and explain its functionality. Suggest improvements for the following aspects of the script:

- The scope of the iterator i is hard-coded to 6. When would this result in problem? How could you choose the iterator in a smarter fashion or change the script logic to avoid this problem?
- Expand the script so that it will also consider other university cities in Bavaria (for illustration use Bamberg, Bayreuth and Augsburg).

(b) Getting listing details

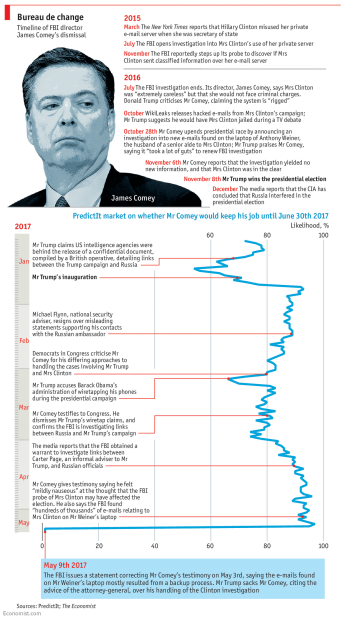
Write a function getApartmentDetails which takes the elements from *allResults* from the above script as input and returns a data frame with listing details. Your script should then extract important information for the listing, at least get the following aspects:

- Rent (Kaltmiete) (Watch out: *gsub(".","","4.000)* yields an empty string, you need to do *gsub("[.]","","4.000)*
- Size (m^2)
- Number of rooms
- District

(c) *Putting it all together

Use *map_df* to run your function for all listings in Würzburg (and potentially other cities). Develop an interesting and appealing set of visualizations which help grasp the apartment market. Think broad and consider coloring, facetting, maps or combinations of different approaches.

2: Comey is going down



As you may have heard James Comey was fired as director of the FBI by US president Donald Trump. The economist had an interesting info graphic which combined news events with stock prices on the prediction market PredictIt.com. You find the stock data in Wuecampus and are asked do recreate a similar plot in ggplot (of course only the lower part).
Do not waste time trying to replicate every detail but try to capture the overall feeling of the plot.

See `http://www.economist.com/blogs/graphicdetail/2017/05/daily-chart-7` for thw full graphic.

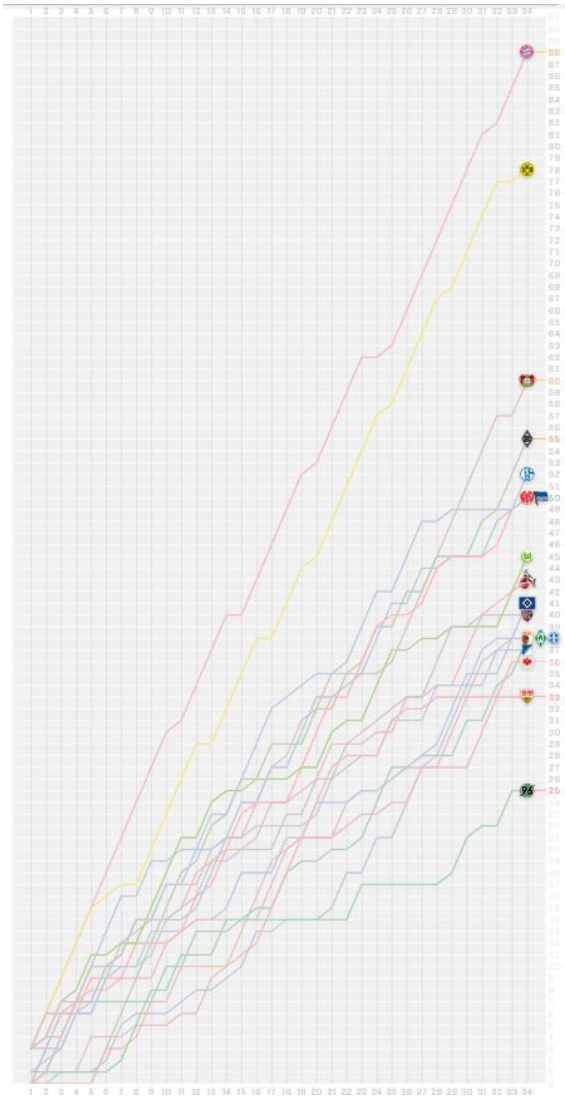
The post `https://stackoverflow.com/questions/25106508/ggplot2-is-there-an-easy-way-to-wrap-annotation-text` may be helpful for labels with word wrap.

3: Zeit Germany Map

The German newspaper Die Zeit has a popular feature where they present an interesting / witty map of Germany. The one on the left presents world heritage sites in Germany. Use wikidata to get the current list (and coordinates) of the world heritage sites in Germany. Subsequently, plot and label them using ggmaps. You do not need to create icons (only if you have to much spare time :)).

See `https://www.r-bloggers.com/gps-basemaps-in-r-using-get_map/` for some alternative ways of styling the map which may help improve the look and feel.

4: Better Bundesliga Tables



Another fairly boring Bundesliga season is finished. The year before the gigantic lead of Bayern and Dortmund prompted the newspaper Zeit to create a "better table:" http://www.zeit.de/sport/2016-04/bundesliga-tabelle-bessere-abstaende
Your task in this assignment is to recreate this chart as closely as possible using ggplot for the new season. The dataset D1 (`http://www.football-data.co.uk/mmz4281/1617/D1.csv`) provides all the necessary information to recreate this point chart. See `http://www.football-data.co.uk/notes.txt` for an exact documentation. For people not interested in sports, a win yields 3 points, a draw 1 point.
Some remarks:

- A good start is to gather the columns *HomeTeam* and *AwayTeam*.
- You do not need to implement the team logos, however nice labelling of the teams (possible abbreviated) next to the final tally should be enough
- Aligning team colors with the line color would be nice. Similarly, the total points on the right are also a nice touch.
- Animating the plot with gganimate (or facetted over matchdays) is the ideal finish.