ETC3250 Lab 7

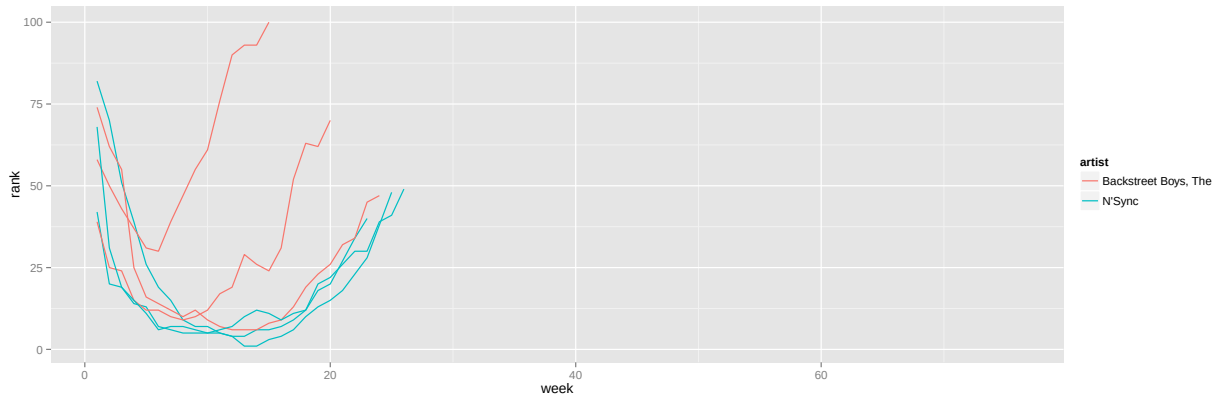
Di Cook

8 September 2015

Making data plots and cleaning data

Task 1

1. Read in the billboard top 100 music data, which contains N'Sync and Backstreet Boys songs that entered the billboard charts in the year 2000
2. Use `tidyr` to convert this data into a long format appropriate for plotting a time series (date on the x axis, chart position on the y axis)
3. Use `ggplot2` to create this time series plot:



Task 2

Read in the flights data:

This dataset contains information on over 300,000 flights that departed from New York City in the year 2013.

1. Using dplyr and the pipe operator, create a data frame consisting of the average arrival delay (`arr_delay`) based on the destination airport (`dest`). Sort this data frame in descending order, so the destination airport with the largest delay is first.
2. Find out the most used airports for each airline carrier.

Task 3

1. Using the `flights` data, create a new column Date using lubridate. You will need to paste together the columns year, month, and day in order to do this. See the `paste` function.
2. Use dplyr to calculate the average departure delay for each date.
3. Plot a time series of the date versus the average departure delay

Assignment 7

On the PISA data from last week. Make a plot to answer this question:

- Does truancy affect math score, on average?
- And pick one other interesting question based on the data dictionary description information to answer using a plot.

Write a couple of sentence explaining why you chose the particular plot design, and how it helps to answer the question.

Turn in two items: a .Rmd document, and the output .pdf from running it. No need to include the R output and plots in your pdf, but the code should be in the Rmd file.