Week 5 Assignment – Tidying and Transforming Data (30 points)

Data storage (10 points)

Create a .csv file containing the information (9 points)
Using “wide” data structure (1 point)
Or, optionally, an SQL DB (extra credit: 1 point)
If DB, normalized (long) data structure (extra point: 1 point)
Implement Database in the Cloud (e.g. on AWS RDS) (extra credit: 1 point)

Importing and Preparing Data (7 points)

Read the information from your .CSV file into R (1 point)
Use tidyr as needed to tidy your data. (3 points)
Use dplyr as needed to transform your data. (3 points)

Reproducibility (4 points)

Data is available on the web (2)
Using R Markdown text and headers (2 points)

Data Analysis (6 points)

Include publication quality graphics or tables (2 points)
Compare the per-city on-time performance for both airlines (1 point)
Compare the overall on-time performance for both airlines (1 point)
Explain the apparent discrepancy between the per-city and overall performances (1 point)
Provide an interesting paradoxical conclusion (1 point)

Documentation (2 points)

Include an overview of your approach, that describes the assigned problem and clearly explains the workflow (1 point)

Provide a conclusion (including any findings and recommendations) (1 point)

Submission (2 points)

Publish to rpubs and provide a link in your assignment submission (1 point)

Publish to GitHub and provide a link in your assignment submission (1 point)