Due Sunday 10/23

Jose

1. Create database connection
2. Loads (upload) raw source csv
3. Extract/download raw csv to R

Now to Thursday

1. ~~Vladimir Add technologies: R, API - tweak~~
2. ETL/ELT, Forecasting, Data Analytics, Visualizations, ML, Logistic Regression, statistics, programmer, big data – Johnny tweak/integrate R/API
3. Sanielle - degree: bachelors, masters, phd ~~and years of experience~~
4. Vlad create 3 (4) tables/dataframe in R

Thursday – Sunday

1. Remove duplicate rows – Sanielle
2. Create integrated R Markdown page
3. Reload clean table to SQL to make them available in the analysis
4. Pull from SQL for analysis calculations
5. Create visualizations
6. Complete the report writeup in RMarkdown

Analysis

Vlad – Salary by Tech Tools > ranked bar chart

Sanielle – Salary by Education, Salary by Simple Job Title, Salary by Skill > ranked bar chart

Jose – proportional Skill by Simple Job Title > mosaic

Johnny – Proportional skills by overall data – stacked bar

Due 10/26 - Presentation

Assign the parts of the Rmarkdown for the pres

Enhance RMarkdown for pres

Presentation Outline

1. Intro – Johnny
   1. Overview Part 1
2. Jose
   1. High Level overview of the dataset
   2. S3 Bucket
   3. AWS
   4. SQL Connection Challenges
3. Vladimir
   1. Cleaning up the raw data
   2. Normalizing the data into multiple tables
   3. Reloading back to SQL
4. Johnny
   1. Process of Analysis
   2. Cleanup needed
5. Sanielle – Analysis
   1. Plots results
   2. Conclusion
      1. By education
      2. By Skill
      3. By Tool
      4. By Job Title