Elizabeth Chabot
PENNY FOR YOUR THOUGHTS
W205.S.2017

Data Output and Visualizations

Detailed Comparison of Tableau and R Shiny in regard to our current problem/RQ.

Problem/RQ

Imagine that a policy group, think tank, research consultancy comes to you and requests that you review the link, if any, between funding and academic success.

Funding can be broken down into multiple areas:

- Federal Funding
- Local Funding
- District Funding
- Donor Funding
- Comparison of Cost of Living by State
- Comparison of Cost of Living by District

As can academic success:

- National Standardized Test Scores
- Grades
- Drop Out Rate
- College Acceptance Rate
- IQ Test

***Note that the above are just suggestions and the team sees many more factors at play for both funding and academic success, but the above were seen as attainable acquisitions during the beginning of the project

For the purpose of our analysis we will review funding at local and federal level, making note that further funding by district is desired, but we were unable to find data for each state. Academic success will be shown by NAEP Math and Reading Scores. We are aware that the current dataset could be enriched in a multitude of ways and that this does not provide a full picture of funding or academic success, but we do believe it's a good starting point and will provide insight on where to focus efforts for improvement and replication.

Speed of Visualization and Dashboard Creation

Elizabeth Chabot
PENNY FOR YOUR THOUGHTS
W205.S.2017

For our project, all data is static and incremental uploads will take place every year to 2 years. We do not need to take into consideration streaming data, high velocity or data approaching 'big data' status. We as a team decided that the best way to output a response to the RQ would be to create a dashboard and story documenting our analysis. The audience we foresee reviewing this analysis will be the public and our stakeholders, lay individuals with little to mild understanding of statistics.

R Shiny:

R Shiny requires a small amount of time to learn how to code if you have experience with R, if you lack experience then it could take months. If working on a team that does not code in R Plotly, Bokeh, Seaborn are other considerations for those who code in Python, while D3 visualizations take the same amount of time if able to create JavaScript, HTML, CSS. What it boils down to is building up a visualization or dashboard from code and packages. It requires one to know enough in one of the above languages and statistics to create a dashboard or visualization. Additional time and effort comes from coding each visualization and having to establish the composition of the dashboard with care.

Our team spent 48 hours working with both Tableau and R Shiny to create a base dashboard and visualizations to see what would be the best option for this current project. In R Shiny we were able to create a base functional dashboard with one visualization.

Tableau:

Tableau for our team provided an easier solution for the creation of visualizations and a dashboard. We were able to create two dashboards and over ten visualizations. Tableau won in speed of creation over R Shiny. Additionally, in a real world setting we would be able to create new visuals quickly and easily to present to stakeholders. R Shiny would be able to create more detailed visuals with more options for statistical analysis. If we were conducting this in RW we would take into consideration in more detail what our stakeholders were requesting and if the analysis required more advanced statistics we would most likely move to R Shiny or create an additional layer (possibly presentation layer) in R before outputting to Tableau.

Maintenance

R Shiny:

Open Source requires constant updates on an as needed basis of packages. This may require for someone to constantly monitor and update the code for the dashboard and your analysis.

Tableau:

Unless we wanted to upgrade we would only to update when necessary aka something being deprecated

Flexibility

R Shiny:

More options for making visuals and bringing in multiple sources, but building up and making changes is harder

Tableau:

Limited to what the software is able to do, but very easy to build and change visualizations on the fly. We did not foresee the creation of overly complex visualizations of the data

Permissions and Access

R Shiny:

Open Source and free, but if what to create permissions and access

Tableau:

Easy to add and take away – LOOK INTO THIS

Cost

R Shiny:

*Free open source version
*Upgrade to R Studio Shiny Server Pro if require more concurrent users to access dashboard, require R Studio support, improved tuning and scaling Cost: Starts at $9,995, as you add users jumps to $14,990-$24,990 large concurrent users will increase pricing

Tableau:

*Currently using free Tableau Student and Tableau Public version
*Upgrade to Tableau Server or Tableau Desktop would require additional cost. Our team foresees Tableau Server being the best option if permissions, higher volume or data and traffic require us to move on from Tableau Public. Cost per Tableau Desktop: $1,500 -

2,500 depending on number of licenses, Cost Monthly for Tableau Server: $300 – Dependent on use-case

Advanced Analytics

R Shiny:

Able to create more advanced analytics

Tableau:

Not quite as advanced, but getting there

Traffic

R Shiny:

The free version of R Shiny has a limitation of 20 users

Tableau:

Tableau Public has no restriction on number of users

Automation

R Shiny:

R is a programming language and thus the majority of the steps required for automation are documented in the process of creation. If the file format changes or new variables are added the code will need to be updated.

Tableau:

Tableau can document the steps of what will need to be done to the data through calculations, parameters and groups, BUT lacks the abilities of R to cleanse and data mine.

Summary

We chose Tableau over R Shiny due to time constraints, not knowing if the visualizations we made would need to be completely reworked and due to the fact that we were able to find a

number of DQ issues through our EDA with Tableau and R setting us up to already have working visualizations in Tableau and only a few rough visualizations in R Studio/R Shiny.

Visualizations

We enriched our visualizations with proficiency and percentile rankings for both reading and math scores, US Regions and US districts. Additionally, we filtered and excluded fields and instances with poor data quality within Tableau and then implemented in our Postgres tables. Metrics to note created in Tableau/Postgres: Aggregated Math and Reading Score, Correlation between Math, Reading, Local Rev and Total Rev, Totals and Averages for all numerical data. Useful Sources:

1. http://nandeshwar.info/data-science-2/tableau-vs-r/
2. https://www.linkedin.com/pulse/r-shiny-v-tableau-dawn-graphics-anand-gupta
3. http://datascience-enthusiast.com/R/R_shiny_Tableau_treemap.html