Stephanie Byng

Wenshan Cao

Gabriel Cruz

Yuan Fang

Virinchi Juttukonda

*Project Proposal*

INFM 600

October 3, 2019

**Team Project - Maryland Socioeconomic and Education Data by County**

*License* - Public Domain

Maryland’s Open Data Portal is a repository containing wide-ranging data from acres of cover crop to youth homicide. Data are conveniently grouped by county, city, or other jurisdiction. We are specifically interested in population and education data, because anecdotally, people talk about how different Montgomery county is compared to other areas in the state. We want to see if this is really the case. How much do Maryland counties vary in terms of demographics, resources, and other attributes? Is Montgomery county a hot spot of income and resources? In this project, we will specifically look for themes related to education that could be used by State and County governments to inform where resources and services could have the biggest impact, by comparing and contrasting the highest and lowest performing counties.

**Questions this data set might help answer**

Prescriptive

* What, if any, are the differences between the highest and lowest performing counties in terms of High School Attainment? Can we identify any trends/correlations?
* Is there a relationship between the cost of living, average income, poverty rate, and/or the amount of money the county spends on education per student?
* How does the state of Maryland compare to the rest of the country in terms of educational outcomes, education costs, etc.?

Predictive

* Does an increase in public school expenditures correlate to an increase in graduation rates, decrease in unemployment and/or a higher average income? Is this predictable - as in increasing expenditures by x% would raise attainment y%?
* Based on projected high school enrollment rates how many more graduates will we see 10 years from now?
* Based on known graduation rates for 2010-2017, can we predict graduation rates by race 10 years from now?

**Risks, Assumptions and Limitations**

For these data sets, we are assuming that the curators of the data (various Maryland Departments) did appropriate data cleansing and normalization of the raw data. Study designs were not specified, therefore we have no knowledge of how the data was collected and processed. This is a risk area as the data could be challenged for validity. For instance, we do not know if data was collected through surveys and/or if the responses are representative of the population. We also do not know how or if the data authors assured that traditionally underrepresented populations were included.

**Preprocessing**

These data provide an opportunity to consolidate multiple data sets together and do data normalization with subject matter that is accessible to all. We will clean the data by removing redundancies from each resource to create a unique data set (data reduction). We will also review each attribute to scan for any outliers or omissions that could have a distorting effect on the results. As needed, we will modify the scale of attributes such that they can be appropriately visualized.

**Analysis**

Data analysis requires running and interpreting statistical queries and presenting results such that laypersons can comprehend. We intend to employ two basic methods.

Basic methods:

* Qualitative and Quantitative Variables
  1. Qualitative Variables describe data that fits into categories such as gender, age, race, etc.
  2. Quantitative Variables are measured in terms of numbers such as population, income, cost of living, etc.
* Descriptive statistics
  1. Examples in mean, variance, standard deviation, etc.

Software for analyzing and calculation

* Microsoft Excel
* R Statistical Programming Language
* Tableau Prep/Desktop

**Hypotheses**

We hypothesize that...

* + ...counties with better high school attainment also have a lower poverty rate, higher income, and/or more residents with college degrees.
  + ...counties that spend more money on education will also report better educational outcomes.
  + ...counties with a higher median income and/or quality of life score will spend more money on education.
  + ...Montgomery county will be in the top 10% in terms of income, educational outcomes and
  + ...Maryland will be in the top 25% of the country in terms of higher education degrees attained.
  + ...regardless of educational spending, black and hispanic students will have a lower rate of graduation than their white or asian counterparts.

**Project Roles**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Task** | **Duties** | **Lead** | **Assist** | **QA/QC** | **Draft Deadline for QA/QC** | **Final Deadline** |
| Data Planning, Documentation, and Metadata | Data dictionary, defining initial needs for processing and analysis. | Stephanie | Yuan | Gabriel | ASAP | ASAP |
| Preprocessing | Cleaning and merging data. (Tableau Prep) | Gabriel | Wenshan | Steph | Oct 18 | Oct 22 |
| Statistical Analysis | Creating dashboards and visualizations needed for presentation. Prescriptive and predictive analysis (Tableau Desktop) | Yuan | Gabriel | Virinchi | Nov 4 | Nov 8 |
| Stakeholder Analysis | Translating stats into knowledge for our designated stakeholders. What does the data mean? How does the info viz answer our questions? This content will make up a big portion of the presentation. | Wenshan | Yuan | Virinchi | Nov 18 | Nov 22 |
| Presentation and Speech | Putting all of the above together into a single slide deck. Writing speech notes. | Stephanie | Virinchi | Wenshan | Nov 29 | Dec 3 |
| Speakers | Presenting to class | Virinchi | Gabriel | Yuan (not QA, but a 3rd speaker) | n/a | Dec 5 |

**Resources**

Maryland Department of Commerce. (2019). *Choose Maryland: Compare Counties - Education.* Retrieved from <https://opendata.maryland.gov/Education/Choose-Maryland-Compare-Counties-Education/63pe-mygy>

Maryland Department of Commerce. (2019). *Choose Maryland: Compare Counties - Quality of Life.* Retrieved from <https://opendata.maryland.gov/Housing/Choose-Maryland-Compare-Counties-Quality-Of-Life/dyym-bjv4>

Maryland Department of Commerce. (2019). *Choose Maryland: Compare States - Education.* Retrieved from <https://opendata.maryland.gov/Education/Choose-Maryland-Compare-States-Education/3bkz-cttp>

Maryland Department of Health. (2019). *SHIP High School Graduation Rate 2010-2017.* Retrieved from <https://opendata.maryland.gov/Health-and-Human-Services/SHIP-High-School-Graduation-Rate-2010-2017/c6t5-8ixv>

Maryland Department of Planning. (2019). *Maryland Counties Socioeconomic Characteristics.* Retrieved from <https://opendata.maryland.gov/Demographic/Maryland-Counties-Socioeconomic-Characteristics/is7h-kp6x>

Maryland State Data Center. (2019). *Public High School Enrollment (Grades 9-12) by Jurisdiction, Actual (2018) and Projected (2019-2028)*. Retrieved from <https://planning.maryland.gov/MSDC/Documents/school_enrollment/school_2019/Table12.xlsx>

**Resources to validate our conclusions related to graduation rates:**

Validation that graduation rates remain fairly stable:

<https://montgomeryschoolsmd.org/departments/sharedaccountability/reports/2019/MSDE_GradRate%20Data_Ann%20Targets.pdf>

Maryland State Department of Education. (2019).*Cohort Graduation Rate Trend Data (2015 - 2018)*. Retrieved from <https://reportcard.msde.maryland.gov/Graphs/#/Graduation/GradRate/1/6/3/1/99/XXXX>