Reports

To make it easier, I have written a report consisting of Introduction, Part1 and Part2.

Below, you can find the pdf's for these reports:

Project Overview (Abstract)

Part 1 Overview

Part 2 Overview

May 19, 2017 / Leave a comment

ZeroR Update

I've calculated ZeroR's again.

You can find the code in GitHub.

May 19, 2017 / Leave a comment

Equal Weight DTCs

I've constructed DTC's with equal weights. For that I used these 2 approaches:

1. I've randomly selected the "No" labels as equal quantitatively as "Yes".
2. I've applied k-means to "No" labels with k = # of "Yes" labeled, and then took the centroids of all k clusters.

Then I used the same DTC classifier and constructed DTCs.

You can find the code on GitHub.

May 14, 2017 / Leave a comment

Plan for last phase

Up to now, I have completed the following steps in part 2:

- Linear Regression to predict the exact grade
- Decision Tree, SVM, Logistic Regression to predict the grade in 5 groups
- Decision Tree, Logistic Regression to predict Top/Bottom 10%-20%

Now, I'll do the following step:

- The classification for top or bottom 10-20% give bad results since there are lots of no's and a few yes's. With standard weights, there are lots of false negatives. Therefore, my next step is to modify weights so that they have equal weight. By doing so, I'll smooth the false positive and false negative numbers.

May 12, 2017 / Leave a comment

Logistic Regressions for Percentages

I've applied Logistic Regressions to predict:

- Top 10% or Not
- Top 20% or Not
- Bottom 10% or Not
- Bottom 20% or Not

Code: MatPercentLogisticRegression.pdf PorPercentLogisticRegression.pdf

May 7, 2017 / Leave a comment

Decision Trees for Percentages

I've constructed decision trees to predict:

- Top 10% or Not
- Top 20% or Not
- Bottom 10% or Not
- Bottom 20% or Not

Code: MatPercentDecisionTree.pdf PorPercentDecisionTree.pdf

May 7, 2017 / Leave a comment

Processing Data for Percentages

I've sorted the data and identified Top 10%, Top 20%, Bottom 10% and Bottom 20%.

In the next step, I will try to predict these without any test scores.

Code: Percentages.pdf

May 7, 2017 / Leave a comment

Search ...

RECENT POSTS

- Reports
  May 19, 2017
- ZeroR Update
  May 19, 2017
- Equal Weight DTCs
  May 14, 2017
- Plan for last phase
  May 12, 2017
- Logistic Regressions for Percentages
  May 7, 2017
- Decision Trees for Percentages
  May 7, 2017
- Processing Data for Percentages
  May 7, 2017
- Confusion Matrices
  May 3, 2017
- Comparison of Different Models
  April 29, 2017
- Logistic Regressions
  April 29, 2017
- SVMs
  April 29, 2017
- Decision Trees
  April 29, 2017
- Linear Regressions
  April 29, 2017
- Getting and Processing Data
  April 29, 2017
- GitHub Integration
  April 26, 2017
- Part 1 Overview
  April 24, 2017
- Planning for the rest of the project
  April 24, 2017

- Regression (New)
  April 14, 2017
- Cluster
  April 13, 2017
- Regressions and Some Plots
  April 9, 2017
- Minimized Data
  April 8, 2017
- Correlation Matrix & Scatter Plot Matrix
  March 24, 2017
- Filtered Missing Data
  March 24, 2017
- Merged Data
  March 23, 2017
- Code for Histograms
  March 18, 2017
- Histograms
  March 18, 2017
- Data Downloaded
  March 18, 2017
- Project Proposal (v3)
  March 18, 2017
- Project Proposal (v2)
  March 2, 2017
- Project Proposal
  March 1, 2017

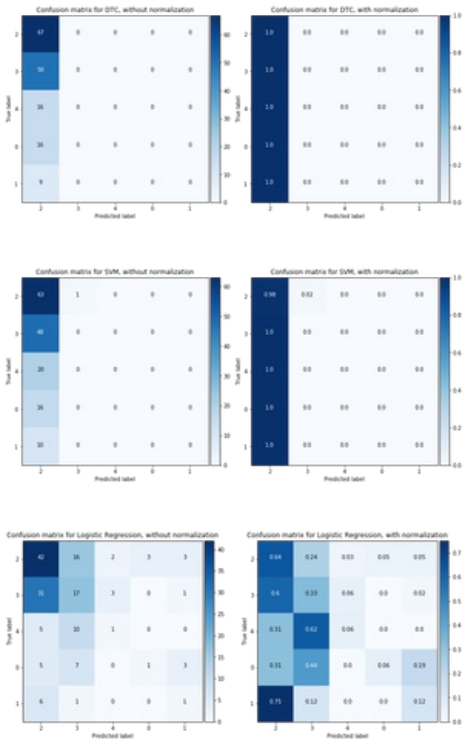
Confusion Matrices

There are the confusion matrices for all cases.

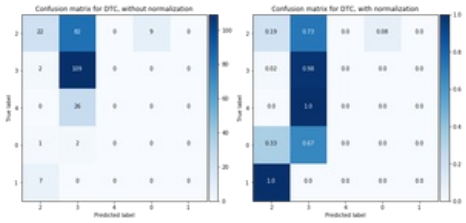
(Note that the numbers may be differ from cross-validated numbers.)

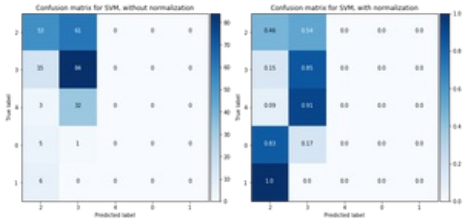
For no test scores given:

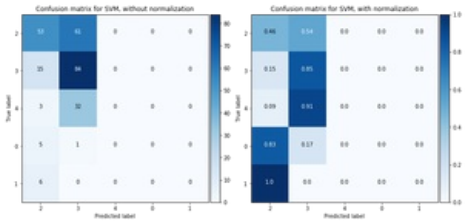
for Mat:



for Por:

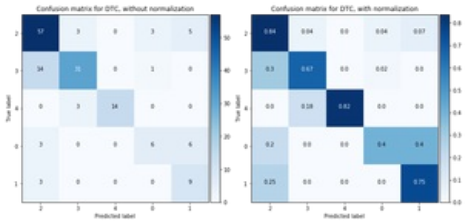


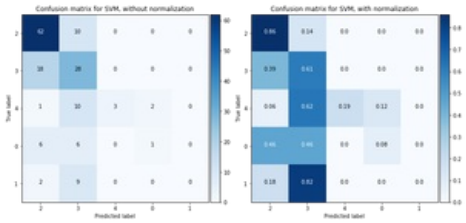


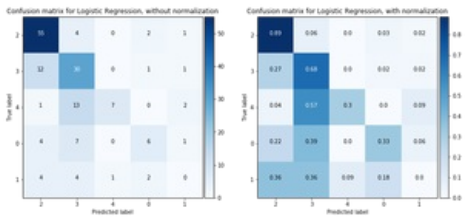


For both test scores given:

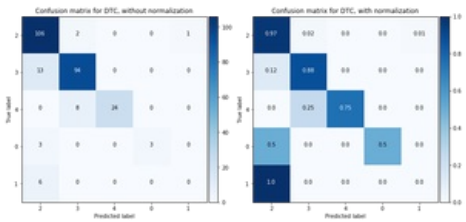
for Mat:

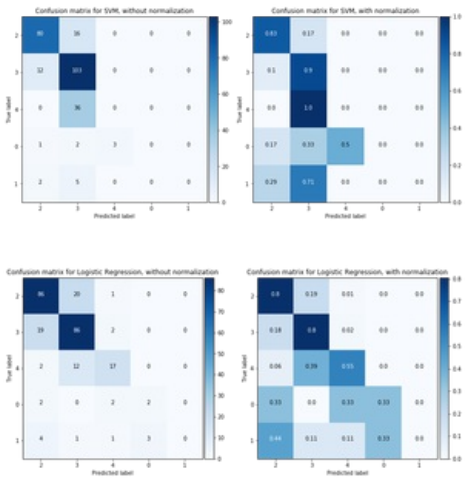






for Por:





May 3, 2017 / Leave a comment

Comparison of Different Models

DBelow, you can find the comparison of OneR, Decision Tree, SVM, Logistic Regression and Linear Regression for my data:

None Given	One R	Decision Tree	SVM	Logistic Regression		Linear Regression
Mat	0.41	0.41	0.42	0.39		—
Por	0.49	0.52	0.53	0.55		—

Both Given	One R	Decision Tree	SVM	Logistic Regression		Linear Regression
Mat	0.63	0.78	0.60	0.62		0.82
Por	0.75	0.85	0.77	0.69		0.85

April 29, 2017 / Leave a comment

Logistic Regressions

I've classified the data using Logistic Regression.

Below are the accuracies:

for Mat:

No test scores: 0.39

Both test scores: 0.62

for Por:

No test scores: 0.55

Both test scores: 0.69

You can find the confusion matrices in the code.

Codes: MatLogisticRegression.pdf PorLogisticRegression.pdf

April 29, 2017 / Leave a comment

SVMs

I've classified the data using SVM.

Below are the accuracies:

for Mat:

No test scores: 0.42

Both test scores: 0.60

for Por:

No test scores: 0.53

Both test scores: 0.77

You can find the confusion matrices in the code.

Codes: MatSVM.pdf PorSVM.pdf

April 29, 2017 / Leave a comment

Decision Trees

I've constructed decision trees for both no test scores given and both test scores are given.

The results are as follows:

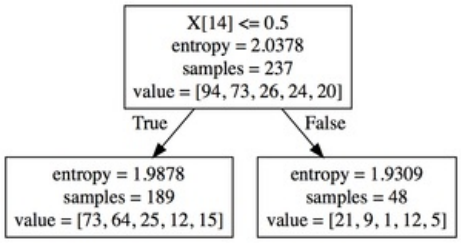
No test scores are given:

for Mat:

Accuracy: 0.41

Best: MaxDepth = 1 (means one R)

Decision Tree:

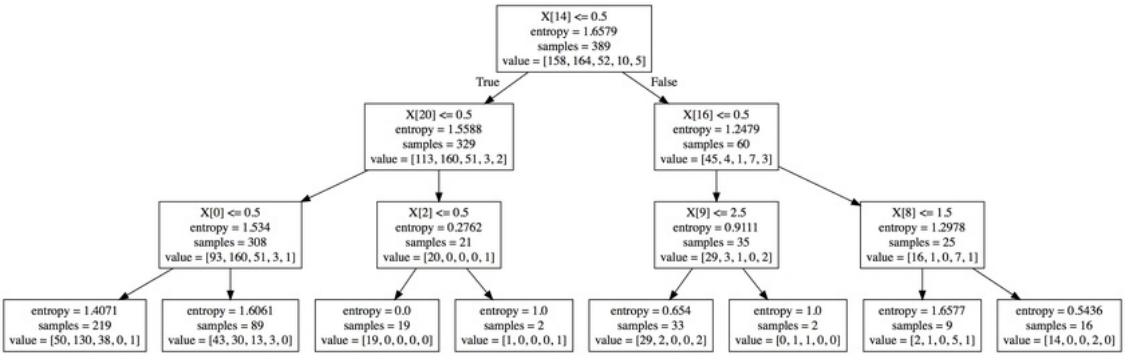


for Por:

Accuracy: 0.52

Best: MaxDepth = 3

Decision Tree:



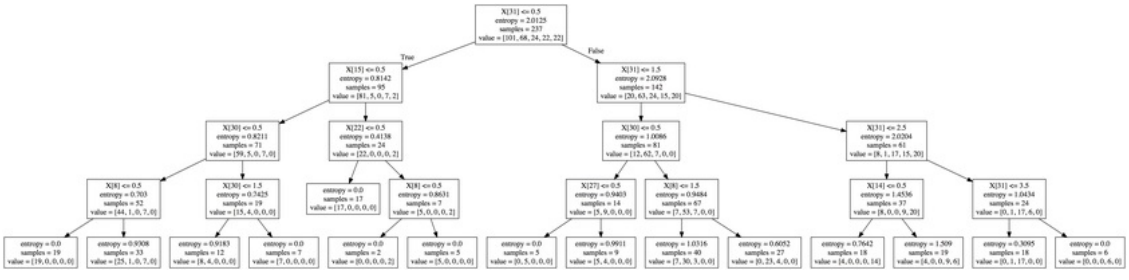
Both test scores are given:

for Mat:

Accuracy: 0.78

Best: MaxDepth = 4

Decision Tree:

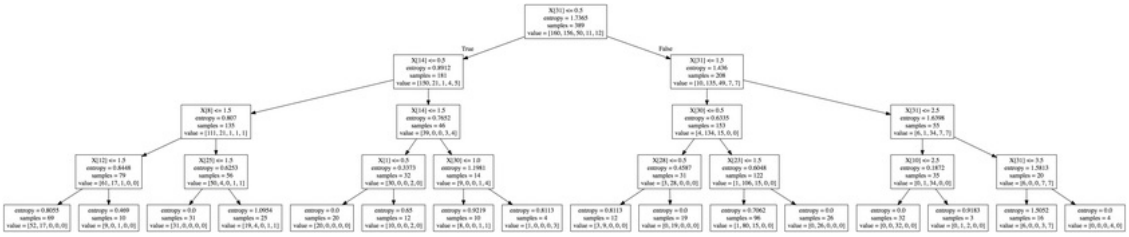


for Por:

Accuracy: 0.85

Best: MaxDepth = 4

Decision Tree:



You can find confusion matrices in the code.pdf.

Code: MatDecisionTree.pdf PorDecisionTree.pdf

April 29, 2017 / Leave a comment

Linear Regressions

I've performed linear regression on (numerical) data.

For Mat, R-squared: 0.822

For Por, R-squared: 0.848

They are very good results and mean that we can very accurately estimate G3 given G1 and G2.

Code (for Mat): MatLinearRegression.pdf

Code (for Por): PorLinearRegression.pdf

April 29, 2017 / Leave a comment

Getting and Processing Data

I've downloaded the data and convert some numeric data to categorical through binning.

I've binned the numerical data as follows:

- age:
  - age <= 18
  - age > 18
- absences
  - absences <= 20
  - 20 < absences <= 40

- 40 < absences <= 60
- 60 < absences <= 80
- 80 < absences

- G1, G2, G3
  - 0 <= score < 4
  - 4 <= score < 8
  - 8 <= score < 12
  - 12 <= score < 16
  - 16 <= score <= 20

Code: CategorizeData.pdf

April 29, 2017 / Leave a comment

Search ...

RECENT POSTS

- Reports
  May 19, 2017
- ZeroR Update
  May 19, 2017
- Equal Weight DTCs
  May 14, 2017
- Plan for last phase
  May 12, 2017
- Logistic Regressions for Percentages
  May 7, 2017
- Decision Trees for Percentages
  May 7, 2017
- Processing Data for Percentages
  May 7, 2017
- Confusion Matrices
  May 3, 2017
- Comparison of Different Models
  April 29, 2017
- Logistic Regressions
  April 29, 2017
- SVMs
  April 29, 2017
- Decision Trees
  April 29, 2017
- Linear Regressions
  April 29, 2017

- Getting and Processing Data
  April 29, 2017
- GitHub Integration
  April 26, 2017
- Part 1 Overview
  April 24, 2017
- Planning for the rest of the project
  April 24, 2017
- Regression (New)
  April 14, 2017
- Cluster
  April 13, 2017
- Regressions and Some Plots
  April 9, 2017
- Minimized Data
  April 8, 2017
- Correlation Matrix & Scatter Plot Matrix
  March 24, 2017
- Filtered Missing Data
  March 24, 2017
- Merged Data
  March 23, 2017
- Code for Histograms
  March 18, 2017
- Histograms
  March 18, 2017
- Data Downloaded
  March 18, 2017
- Project Proposal (v3)
  March 18, 2017
- Project Proposal (v2)
  March 2, 2017
- Project Proposal
  March 1, 2017

GitHub Integration

I've created a git repository for my existing code. You can view all existing and future codes from github.com/berkanteber/cs210project.

I've also created a project page on GitHub. It has still some work to do, but you can reach it from berkanteber.github.io/cs210project.

April 26, 2017 / Leave a comment

Part 1 Overview

(Code: Part1Overview.pdf)

In the first part of the project, I've examined the nation-wide education and R&D data with GDP per capita and GINI index.

I've used the data on education spendings and PISA scores from OECD, and I've used R&D, GDP and GINI index data from World Bank.

Then, I've merged all data as follows:

- for education data, I used the same-year data,
- for PISA scores, I used the latest data,
- for other data, I used the same-year data if possible; then I first looked at the following year's data and then the previous year's data.

Finally, I've eliminated the countries with missing data.

As a result, I've come up with a dataset with 26 countries with the attributes below:

1. Primary to Non-tertiary Education (% of GDP)
2. Public Tertiary Education (% of GDP)
3. Private Tertiary Education (% of GDP)
4. PISA Scores (Reading)
5. PISA Scores (Math)
6. PISA Scores (Science)
7. R&D Expenditures (% of GDP)
8. Researchers in R&D (per million people)
9. High Technology Exports (% of manufactured exports)

10. Resident Patent Numbers (per 100 people)
11. Non-resident Patent Numbers (per 100 people)
12. GDP per capita
13. GINI index

Then, I've applied linear regression on these attributes and I've reached the following results:

- The spendings on primary to non-tertiary education spendings have no significant effect on PISA scores.
- Public tertiary education spendings and PISA math scores explain 54% of the variation in R&D expenditures.
- Public tertiary education spendings explain 34% of the variation in researchers in R&D.
- Education and R&D data are jointly insignificant to explain the variation in high technology data.
- Private tertiary education spendings and R&D expenditures explain 59% of the variation in resident patents.
- Private tertiary education spendings and R&D expenditures explain 63% of the variation in nonresident patents.
- Primary to non-tertiary education spendings, private tertiary education and non-resident patent number explains 58% of the variation in GDP per capita.
- PISA reading and math scores explain 73% of the variation in GINI index.

Some comments:

- The results saying that spendings on primary to non-tertiary education doesn't explain the variation in PISA scores was unexpected for me. Since, we see PISA scores explaining some other variables in other regressions, my interpretation is that PISA scores are not solely related to the spendings but the quality also.
- The result regarding to high technology exports was also unexpected.
- Public tertiary spendings explain the variation in R&D data, however they are insignificant to explain the variation in patent data. My interpretation is that since the variables are not truely independent R&D expenditure also contains information about the public tertiary education spendings.
- We can see that 58% of the variation in GDP per capita is explained by primary to non-tertiary education spendings, private tertiary education spendings and nonresident patent numbers. However, we cannot decide whether high GDP per capita increases the percentage of the education spendings or higher education spendings increase GDP per capita. I am also indecisive about this question. On the other hand, I think, the relation between nonresident patent numbers and GDP per capita is an indicator of brain drain (human capital flight).

April 24, 2017 / Leave a comment

Planning for the rest of the project

So far, I've analysed the data sets from OECD and World Bank about education, R&D, GDP per capita and GINI index. I have performed some linear regressions on these data sets. There have been both some significant and insignificant results.

My next move will be finalizing the interpretations on these data.

Then, I will move on to a new data set from the site below:

https://archive.ics.uci.edu/ml/datasets/Student+Performance#

The reason of that is to analyze the effect of individual differences on education.

Previously, I analyzed the nation-wide education data and it's effects nation-wide.

Now, it's time to focus on a more individualistic study.

To do so, I will have been analyzed education on wider aspects.

April 24, 2017 / Leave a comment

Regression (New)

I performed the regressions using statsmodels this time.
Since I have test statistics, I performed inference more accurately.

Below, you can find the resulting regressions for each possible output. I have chosen the least restrictive regression according to f-test. You can find the intermediary steps are intermediary steps in the code.

Regressions:

• pryntry on pisa_read: non-significant
• pryntry on pisa_math: non-significant
• pryntry on pisa_science: non-significant

• pisa_math and try_public on rd_exp: significat, R-squared = 0.543
• try_public on pisa_researchers_rd: significant, R-squared = 0.342
• pisa_read, pisa_math, pisa_science, pryntry, try_public, try_private, rd_exp, researchers_rd on high_tech_exports: non-significant

• try_private, rd_exp on patent_res_per100people: significant, R-squared = 0.593
• try_private, rd_exp on patent_nres_per100people: significant, R-squared = 0.632

• pryntry, try_private, patent_nres_per100people on gdp_percapita_dividedby1000: significant, R-squared = 0.577
• pisa_read, pisa_math on gini_index: significant, R-squared = 0.725

Code: LinearRegressionNew.pdf

April 14, 2017 / Leave a comment

Cluster

I clustered the countries into 2 groups according to GDP per Capita and GINI index.

When I looked at the data, I figured out that the clusters are almost like the below:

Cluster 1: Developed Countries Cluster 2: Developing Countries

I plan to using other data points such as education, R&D or patent numbers to predict the cluster.

Code: Cluster.pdf

Data: MinimizedDataFullClustered.pdf

April 13, 2017 / Leave a comment

Regressions and Some Plots

(Note: I used sklearn for this regressions. Since there is no test statistics there, I used only R-squared to interpret. I will use statsmodels for more accurate interpretations.)

I performed some linear regressions on the data I have.

Below, you can find X, Y and R-squared for each regression:

1. Primary to Non-Tertiary Education Spendings on PISA Scores
   (R-squared = 0.029) (Plot Available)

   Comment: Spendings on primary to non-tertiary education explain a very little portion of PISA scores.

2. Education Spendings on R&D Expenditures
   (R-squared = 0.080) (Plot Available)

   Comment: Education spendins explain a little portion of R&D expenditures.

3. PISA Scores and Tertiary Education Spendings on R&D Expenditures
   (R-squared = 0.566)

   Comment: If we replace primary to non-tertiary education with PISA scores, R-squared increases significantly and now a large portion of R&D expenditures is explained.

4. PISA Scores and Education Spendings on R&D Expenditures
   (R-squared = 0.571)

   Comment: If we add spendings on primary to non-tertiary education to the previous regression R-squared increases a little bit.

5. Education Spendings on Researchers in R&D
   (R-squared = 0.025) (Plot Available)

   Comment: Education spendins explain a little portion of researchers in R&D.

6. PISA Scores and Tertiary Education Spendings on Researchers in R&D
   (R-squared = 0.138)

   Comment: If we replace primary to non-tertiary education with PISA scores, R-squared increases significantly and now some portion of R&D expenditures is explained but not much.

7. PISA Scores and Education Spendings on Researchers in R&D
   (R-squared = 0.204)

   Comment: If we add spendings on primary to non-tertiary education to the previous regression R-squared increases from 13.8% to 20.4%. It is significant.

8. PISA Scores, Education Spendings, R&D Expenditures and Researchers in R&D on High Technology Exports
   (R-squared = 0.356)

   Comment: Education data and R&D data explains a significant portion of high technology exports.

9. PISA Scores and Education Spendings on High Technology Exports
   (R-squared = 0.336)

   Comment: Only education data explains a little less, however it is still significant.

10. PISA Scores, Education Spendings, R&D Expenditures and Researchers in R&D on Patent Numbers
    (R-squared = 0.685)

    Comment: Education data and R&D data explains a large portion of patent numbers.

11. PISA Scores and Ecucation Spendings on Patent Numbers
    (R-squared = 0.460)

    Comment: Only education data explains a little less, however it is still significant.

12. PISA Scores, Education Spendings, R&D Expenditures and Researchers in R&D on GDP per Capita
    (R-squared = 0.515)

    Comment: Education data and R&D data explains a large portion of GDP per capita.

13. PISA Scores and Education Spendings on GDP per Capita
    (R-squared = 0.341)

Comment: Only education data explains a little less, however it is still significant.

14. PISA Scores and Education Spendings on GINI index
    (R-squared = 0.795)

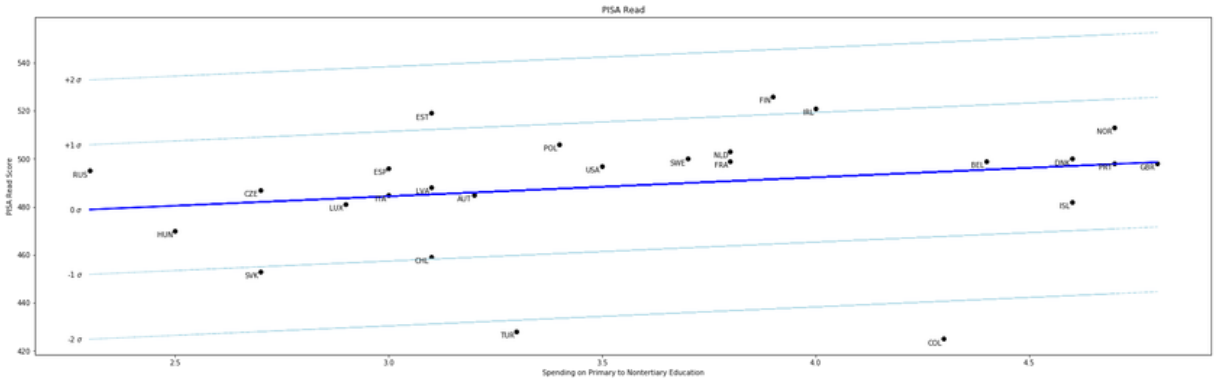
    Comment: Education data explains a very large portion of GINI index.

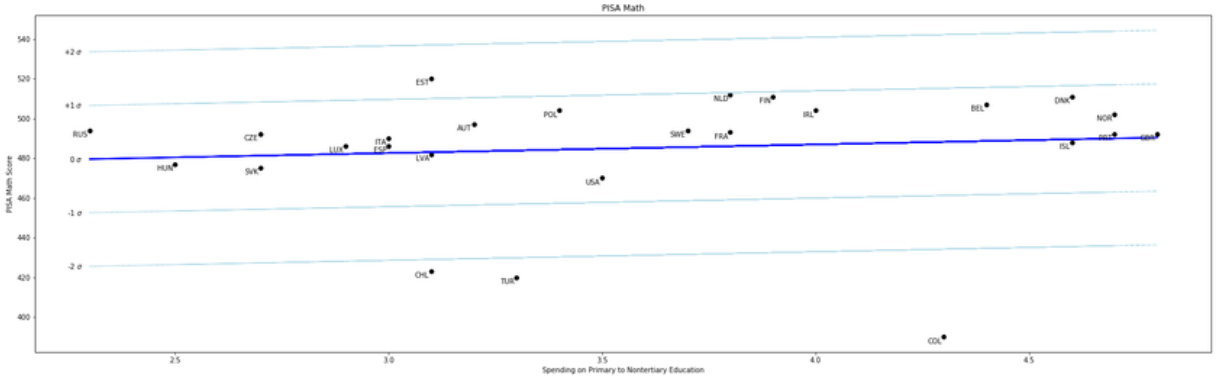
15. Education Spendings on GINI index
    (R-squared = 0.230) (Plot Available)

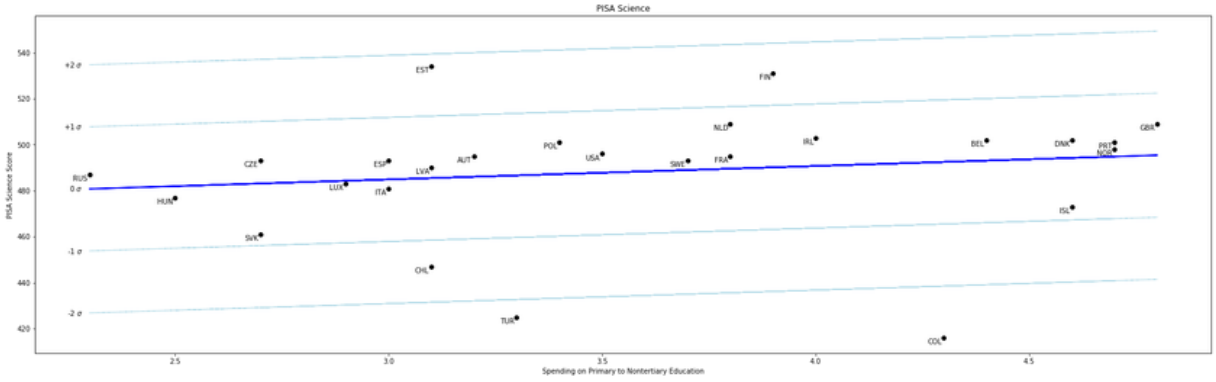
    Comment: Only education spendings explains less, however it is still significant.

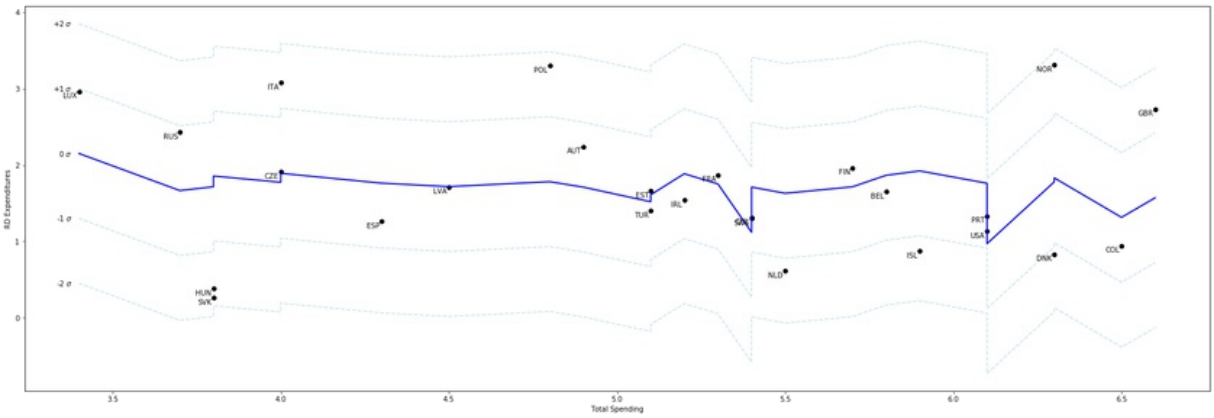
Code: LinearRegression.pdf

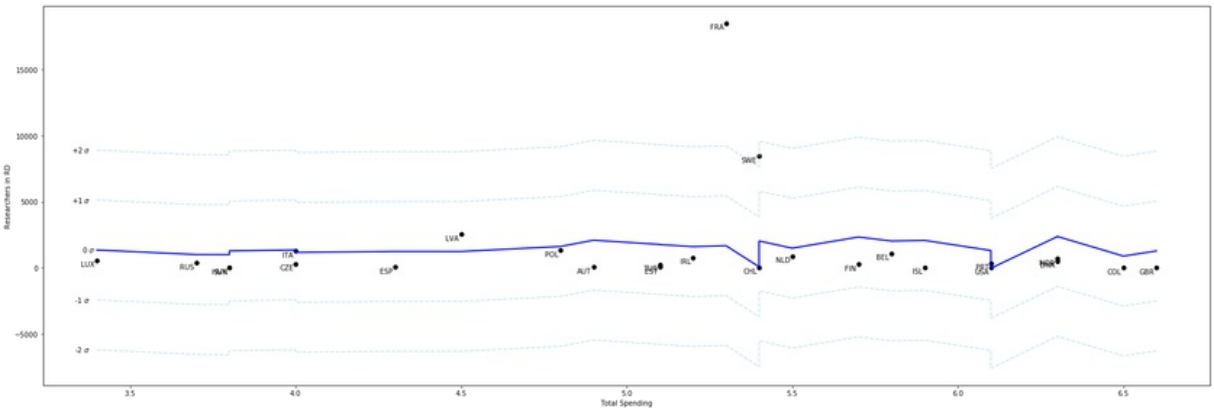
Plots:

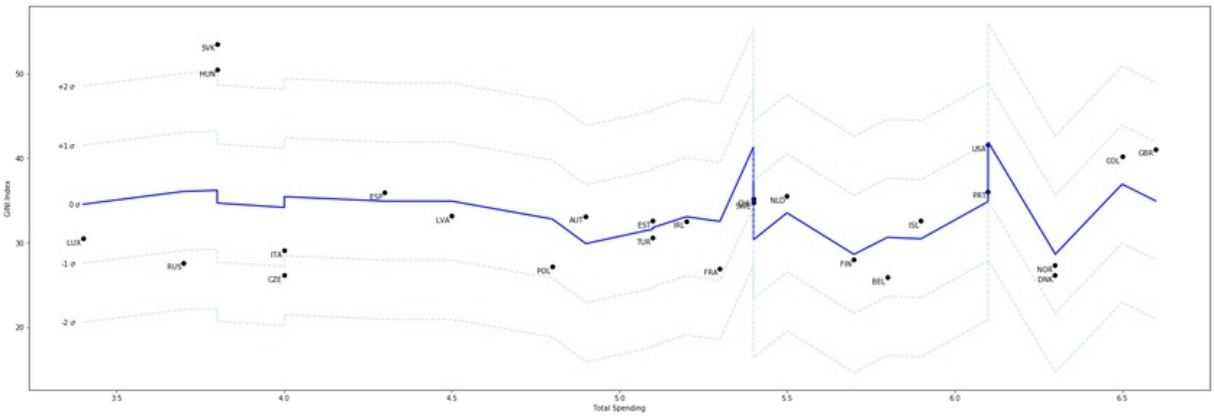












April 9, 2017 / Leave a comment

Minimized Data

I threw 2 data sets since they can be derived from others.

try : try_public + try_private
total: pryntry + try_public + try_private

Code: MinimizeData.pdf

Data: MinimizedDataFull.pdf

April 8, 2017 / Leave a comment

Search …

RECENT POSTS

- Reports
  May 19, 2017
- ZeroR Update
  May 19, 2017
- Equal Weight DTCs
  May 14, 2017
- Plan for last phase
  May 12, 2017
- Logistic Regressions for Percentages
  May 7, 2017
- Decision Trees for Percentages
  May 7, 2017
- Processing Data for Percentages
  May 7, 2017
- Confusion Matrices
  May 3, 2017
- Comparison of Different Models
  April 29, 2017
- Logistic Regressions
  April 29, 2017
- SVMs
  April 29, 2017
- Decision Trees
  April 29, 2017
- Linear Regressions
  April 29, 2017
- Getting and Processing Data
  April 29, 2017
- GitHub Integration
  April 26, 2017
- Part 1 Overview
  April 24, 2017
- Planning for the rest of the project
  April 24, 2017
- Regression (New)
  April 14, 2017
- Cluster
  April 13, 2017
- Regressions and Some Plots
  April 9, 2017
- Minimized Data

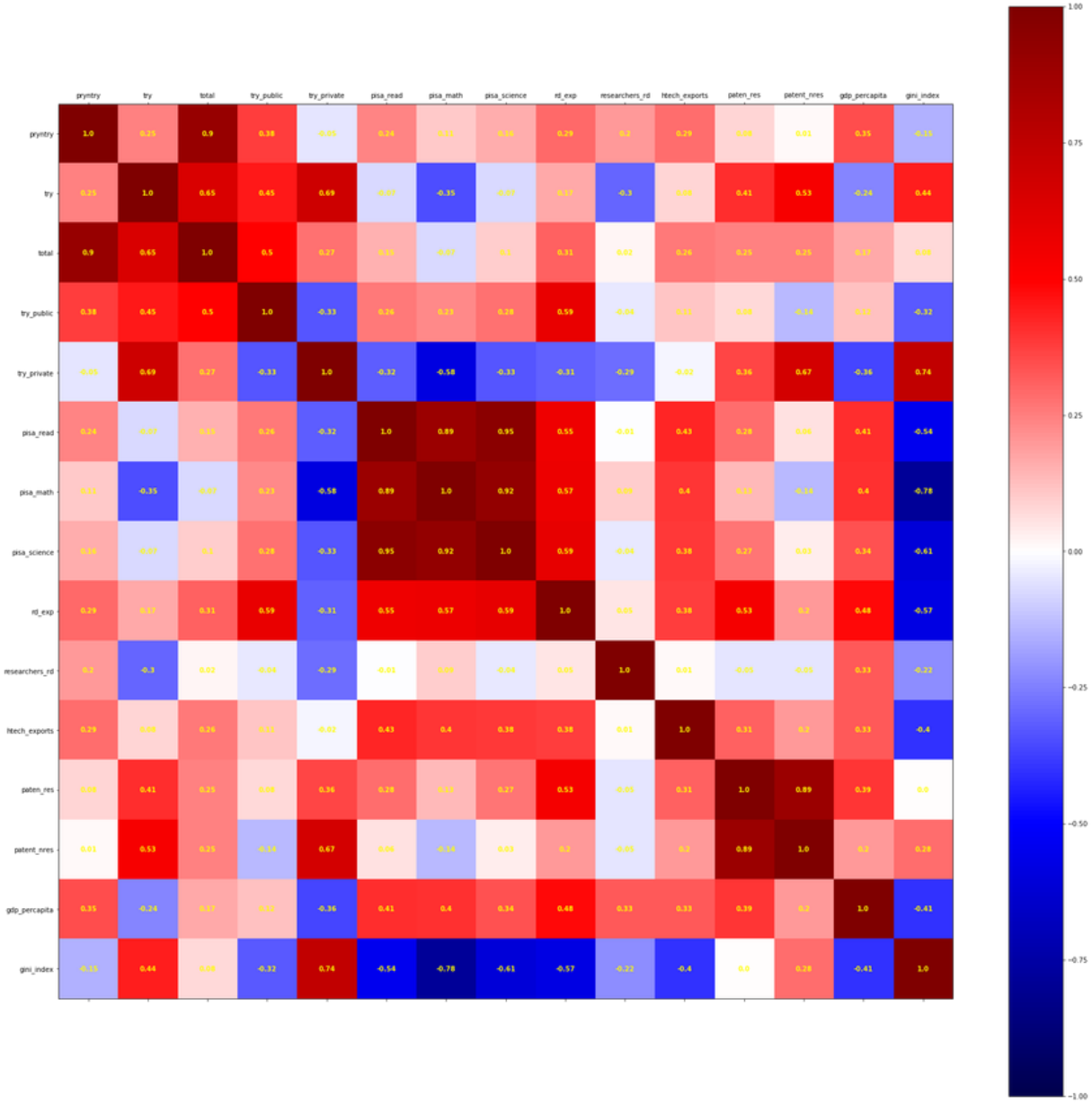
April 8, 2017

- Correlation Matrix & Scatter Plot Matrix
  March 24, 2017
- Filtered Missing Data
  March 24, 2017
- Merged Data
  March 23, 2017
- Code for Histograms
  March 18, 2017
- Histograms
  March 18, 2017
- Data Downloaded
  March 18, 2017
- Project Proposal (v3)
  March 18, 2017
- Project Proposal (v2)
  March 2, 2017
- Project Proposal
  March 1, 2017

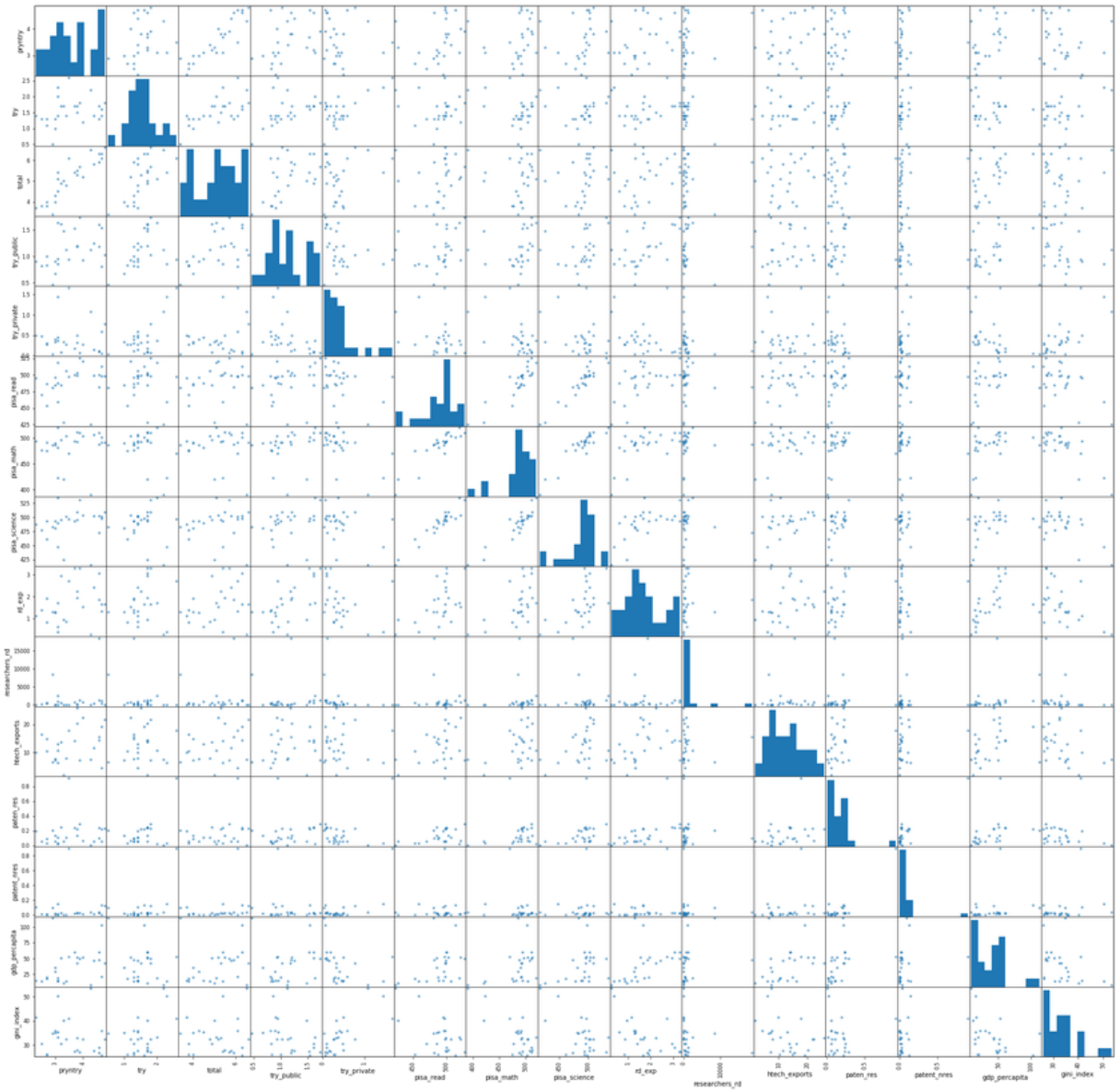
Correlation Matrix & Scatter Plot Matrix

Code : Statistics-Plottings.pdf

Correlation Matrix:



Scatter Plot Matrix:



March 24, 2017 / Leave a comment

Filtered Missing Data

After merging all data, I had 41 countries with their data. However, in some of the countries, I had some missing data. Therefore, I excluded thouse countries.

In final version, I have 26 countries.

Code for filtering: FilterMissingData.pdf

Data in .pdf format: AllDataFull.pdf

March 24, 2017 / Leave a comment

Merged Data

I merged all data.

I excluded public and private spending on education since there were inconsistencies between these 2 data and the other data I have.

In the resulting data I have 41 countries which were originally in OECD data. For World Bank data, I also used only these countries.

While merging,

- for education data, I used the same-year data
- for PISA scores, I used the latest data
- for other data, I used the same-year data if possible; then I first looked at the following year's data and then the previous year's data

In conclusion, I combined all data and exported that to a .csv file. From now on, I will use this data.

Code for merging all data: MergeData.pdf

Data in .pdf format: AllData.pdf

March 23, 2017 / Leave a comment

Code for Histograms

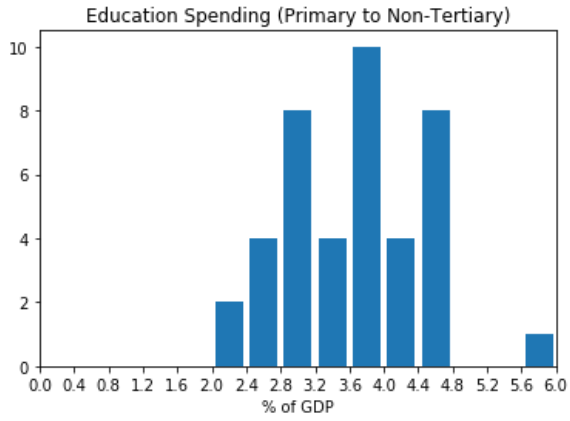
Jupyter notebook in pdf format is below:

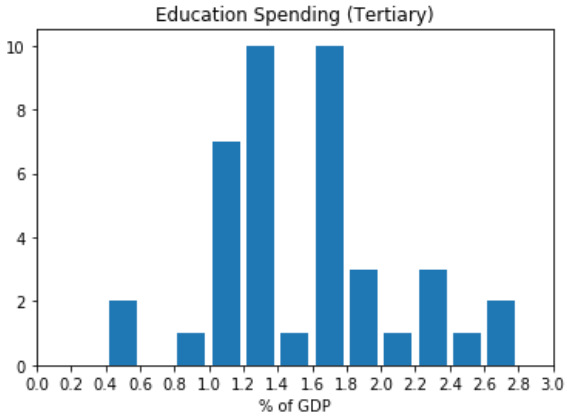
Histograms.pdf

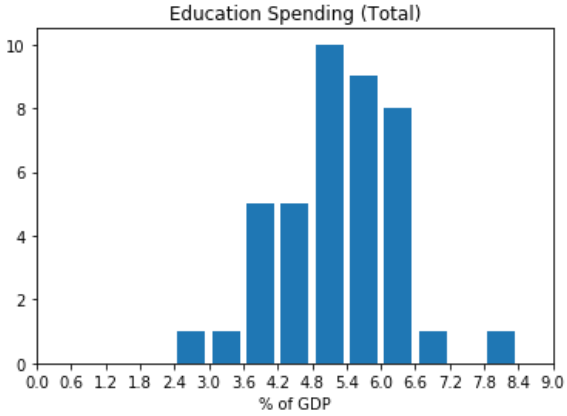
March 18, 2017 / Leave a comment

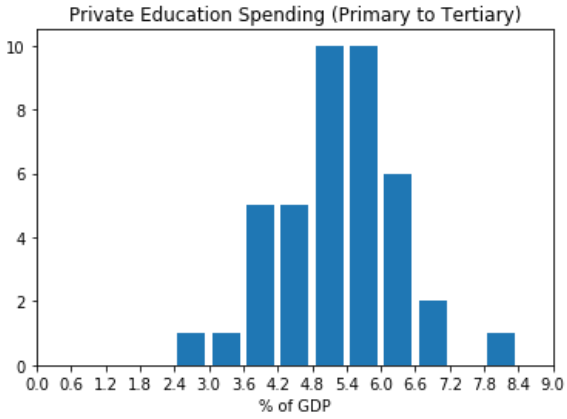
Histograms

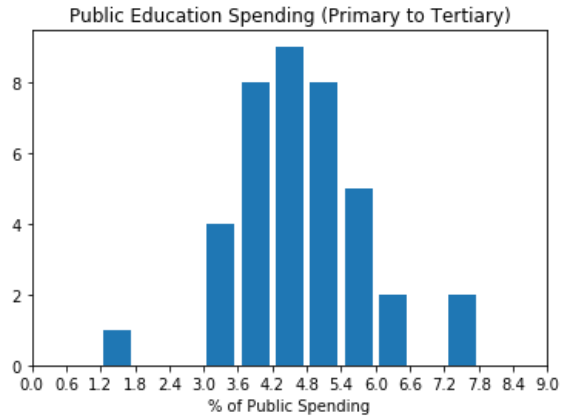
Histograms of each data are below:

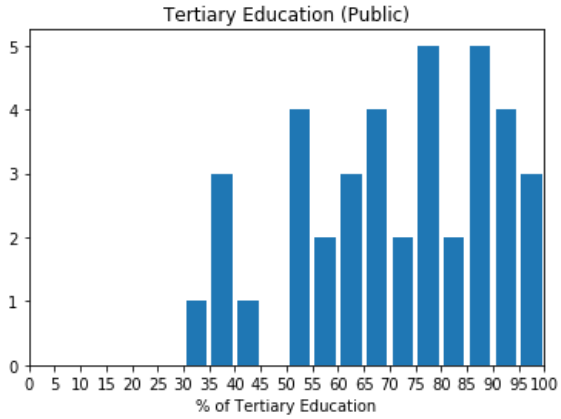


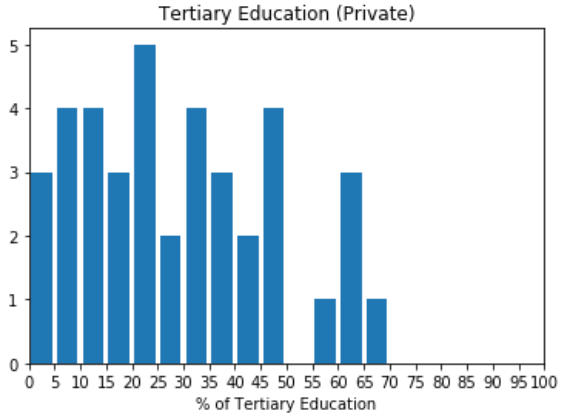

Education Spending (Tertiary)
% of GDP

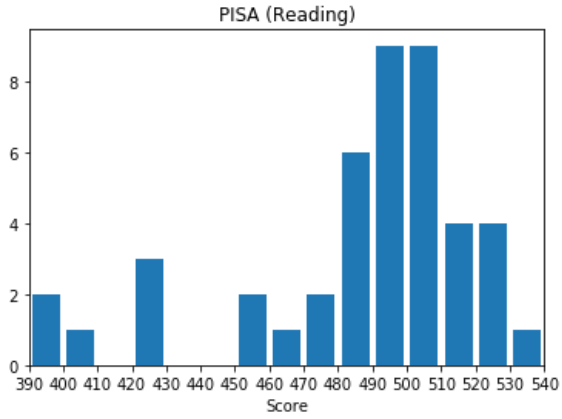

Education Spending (Total)
% of GDP

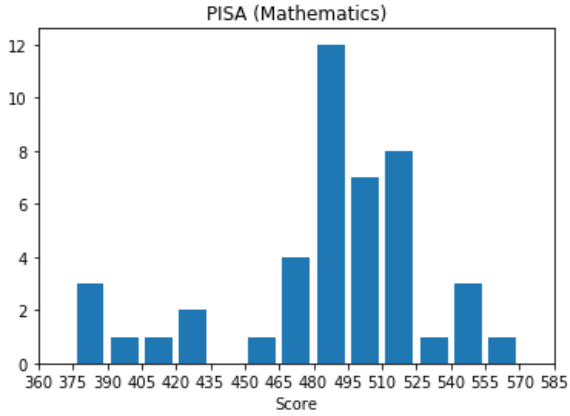

Private Education Spending (Primary to Tertiary)
% of GDP

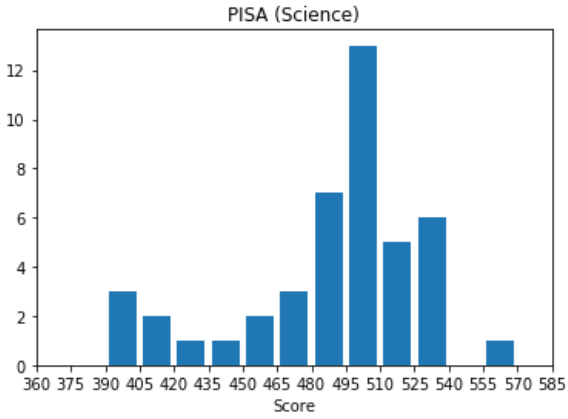

Public Education Spending (Primary to Tertiary)
% of Public Spending

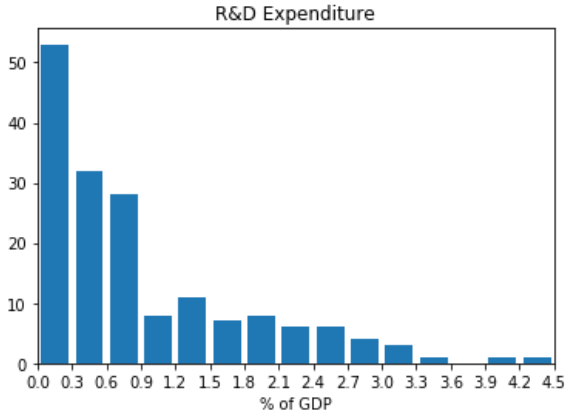

Tertiary Education (Public)
0 1 2 3 4 5
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
% of Tertiary Education

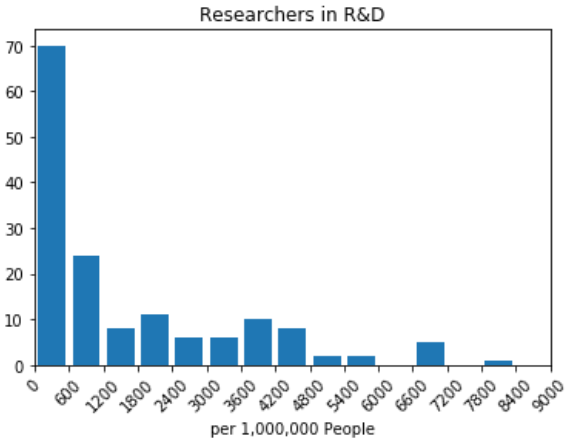

Tertiary Education (Private)
0 1 2 3 4 5
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
% of Tertiary Education

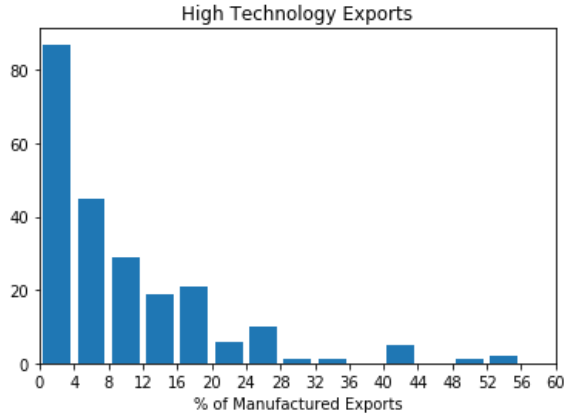

PISA (Reading)
0 2 4 6 8
390 400 410 420 430 440 450 460 470 480 490 500 510 520 530 540
Score

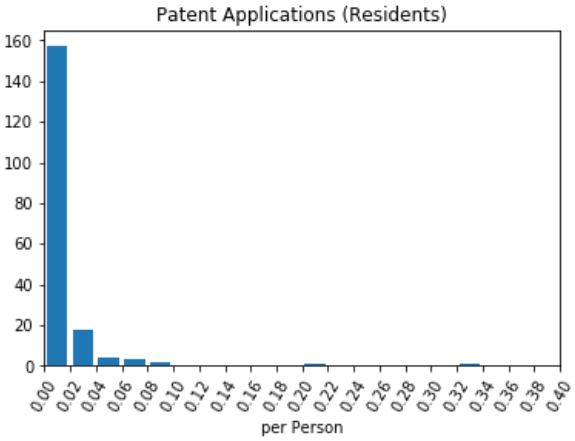

PISA (Mathematics)
0 2 4 6 8 10 12
360 375 390 405 420 435 450 465 480 495 510 525 540 555 570 585
Score

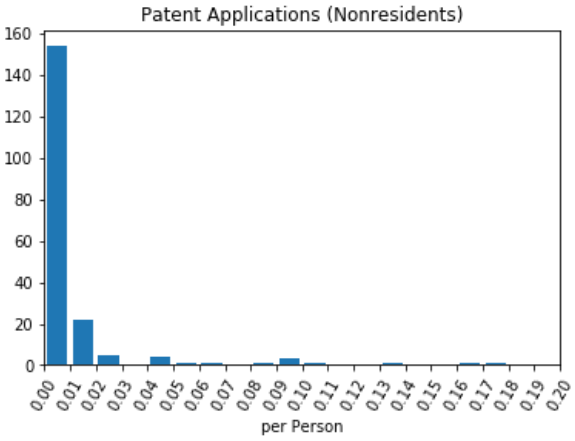

PISA (Science)
0 2 4 6 8 10 12
360 375 390 405 420 435 450 465 480 495 510 525 540 555 570 585
Score

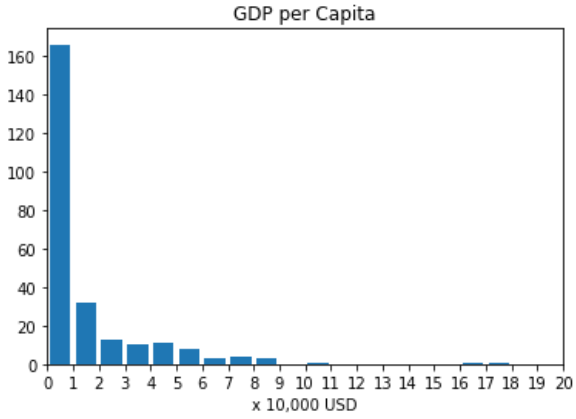

R&D Expenditure
0 10 20 30 40 50
0.0 0.3 0.6 0.9 1.2 1.5 1.8 2.1 2.4 2.7 3.0 3.3 3.6 3.9 4.2 4.5
% of GDP

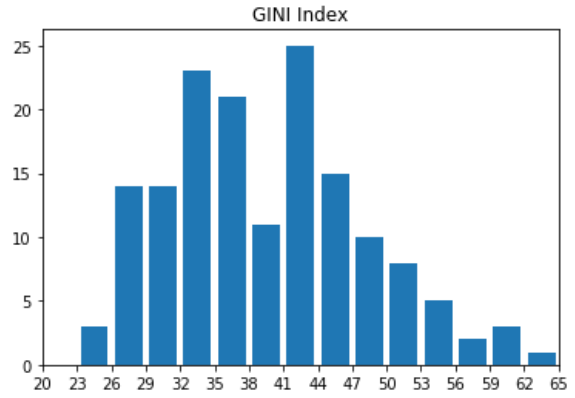

Researchers in R&D
0 10 20 30 40 50 60 70
0 600 1200 1800 2400 3000 3600 4200 4800 5400 6000 6600 7200 7800 8400 9000
per 1,000,000 People


High Technology Exports
0 20 40 60 80
0 4 8 12 16 20 24 28 32 36 40 44 48 52 56 60
% of Manufactured Exports


Patent Applications (Residents)
per Person


Patent Applications (Nonresidents)
per Person


GDP per Capita
x 10,000 USD


GINI Index

Data Downloaded

All data are downloaded in .csv format.

– Data downloaded from OECD contains only last available data.

– Data downloaded from World Bank contains all data.

March 18, 2017 / Leave a comment

Project Proposal (v3)

Performance Analysis of Education According to Various Factors

Education industry is a very large industry according to both size and importance. Governments and/or people spend very large amount of money for this industry, and in return both society and the people who took education benefits from it.

However, there are lots of discussions on the approach to the education. Some people think that the state should provide education to its citizens, whereas others think that private sector can provide better and more efficient service. Also, there is another discussion about the consequences of education. Is it beneficial to the society or to the person more?

In this project, I will analyse the spendings on education both public and private, the distribution of these spendings over education levels (pre-university, university), and the consequences related to the education such as test scores or its effects on society and economy. Then, I will try to find some correlations about all these factors and consequences.

(According to these correlations, I may develop a model to classify countries in different groups such as developed and developing countries according to their spendings on education and some other indicators.)

Below there are links to some data that may be useful on this project:

Education Spending

https://data.oecd.org/eduresource/education-spending.htm

Private Spending on Education:

https://data.oecd.org/eduresource/private-spending-on-education.htm

Public Spending on Education:

https://data.oecd.org/eduresource/public-spending-on-education.htm

Spending on Tertiary Education:

https://data.oecd.org/eduresource/spending-on-tertiary-education.htm

PISA Scores:

https://data.oecd.org/pisa/reading-performance-pisa.htm

https://data.oecd.org/pisa/mathematics-performance-pisa.htm

https://data.oecd.org/pisa/science-performance-pisa.htm

R&D Expenditures:

http://data.worldbank.org/indicator/GB.XPD.RSDV.GD.ZS?end=2013&name_desc=false&start=1996&view=chart

Researchers in R&D:

http://data.worldbank.org/indicator/SP.POP.SCIE.RD.P6?view=chart

High Technology Exports:

http://data.worldbank.org/indicator/TX.VAL.TECH.MF.ZS?view=chart

Patent Applications (Residents):

http://data.worldbank.org/indicator/IP.PAT.RESD?view=chart

Patent Applications (Nonresidents):

http://data.worldbank.org/indicator/IP.PAT.NRES?view=chart

Population:

http://data.worldbank.org/indicator/SP.POP.TOTL?view=chart

GDP Per Capita:

http://data.worldbank.org/indicator/NY.GDP.PCAP.CD

GINI Index:

http://data.worldbank.org/indicator/SI.POV.GINI?view=chart

RECENT POSTS

- Reports
  May 19, 2017
- ZeroR Update
  May 19, 2017
- Equal Weight DTCs
  May 14, 2017
- Plan for last phase
  May 12, 2017
- Logistic Regressions for Percentages
  May 7, 2017
- Decision Trees for Percentages
  May 7, 2017
- Processing Data for Percentages
  May 7, 2017
- Confusion Matrices
  May 3, 2017
- Comparison of Different Models
  April 29, 2017
- Logistic Regressions
  April 29, 2017
- SVMs
  April 29, 2017
- Decision Trees
  April 29, 2017
- Linear Regressions
  April 29, 2017
- Getting and Processing Data
  April 29, 2017
- GitHub Integration
  April 26, 2017
- Part 1 Overview
  April 24, 2017
- Planning for the rest of the project
  April 24, 2017
- Regression (New)
  April 14, 2017
- Cluster
  April 13, 2017
- Regressions and Some Plots
  April 9, 2017
- Minimized Data
  April 8, 2017
- Correlation Matrix & Scatter Plot Matrix
  March 24, 2017
- Filtered Missing Data
  March 24, 2017

- Merged Data
  March 23, 2017
- Code for Histograms
  March 18, 2017
- Histograms
  March 18, 2017
- Data Downloaded
  March 18, 2017
- Project Proposal (v3)
  March 18, 2017
- Project Proposal (v2)
  March 2, 2017
- Project Proposal
  March 1, 2017

Project Proposal (v2)

Performance Analysis of Education According to Various Factors

Education industry is a very large industry according to both size and importance. Governments and/or people spend very large amount of money for this industry, and in return both society and the people who took education benefits from it.

However, there are lots of discussion on the approach to the education. Some people think that the state should provide education to its citizens, whereas others think that private sector can provide better and more efficient service. Also, there is another discussion about the consequences of education. Is it beneficial to the society or to the person more?

In this project, I will analyse the spendings on education both public and private, the distribution of these spendings over education levels (pre-university, university), and the consequences related to the education such as test scores, some prizes or society related topics. Then, I will try to find some correlations about all these factors and consequences.

(According to these correlations, I may develop a model to classify countries in different groups such as developed and developing countries according to their spendings on education and some other indicators.)

Below there are links to some data that may be useful on this project:

Education Spending

https://data.oecd.org/eduresource/education-spending.htm

Private Spending on Education:

https://data.oecd.org/eduresource/private-spending-on-education.htm

Public Spending on Education:

https://data.oecd.org/eduresource/public-spending-on-education.htm

Spending on Tertiary Education:

https://data.oecd.org/eduresource/spending-on-tertiary-education.htm

Main Science and Technology Indicators:

http://stats.oecd.org/viewhtml.aspx?datasetcode=MSTI_PUB&lang=en

PISA Scores:

https://data.oecd.org/pisa/reading-performance-pisa.htm

https://data.oecd.org/pisa/mathematics-performance-pisa.htm

https://data.oecd.org/pisa/science-performance-pisa.htm

Nobel Prize Nomination Archive:

http://www.nobelprize.org/nobel_organizations/nobelmedia/nobelprize_org/developer/

March 2, 2017 / Leave a comment

Project Proposal

Comparison of Public and Private Sector on Education

There is a discussion about the role of the government on different sectors: health, education, social security etc. There are several viewpoints for this question like ethics, efficiency, performance and feasibility.

In this project I will consider efficiency and performance of these both sectors. I will analyse the public and private spendings in different countries and compare them to the indicators of the performance like test results or number of prize winners. In the end of the project, I will try to find out the best approach in education according to efficiency and performance.

Below there are links to some data that may be useful on this project:

Private Spending on Education:

https://data.oecd.org/eduresource/private-spending-on-education.htm

Public Spending on Education:

https://data.oecd.org/eduresource/public-spending-on-education.htm

PISA Scores:

https://data.oecd.org/pisa/reading-performance-pisa.htm

https://data.oecd.org/pisa/mathematics-performance-pisa.htm

https://data.oecd.org/pisa/science-performance-pisa.htm

Nobel Prize Nomination Archive:

http://www.nobelprize.org/nobel_organizations/nobelmedia/nobelprize_org/developer/

March 1, 2017 / Leave a comment

PAGE 5

Search ...

RECENT POSTS

- Reports
  May 19, 2017
- ZeroR Update
  May 19, 2017
- Equal Weight DTCs
  May 14, 2017
- Plan for last phase
  May 12, 2017
- Logistic Regressions for Percentages
  May 7, 2017
- Decision Trees for Percentages
  May 7, 2017
- Processing Data for Percentages
  May 7, 2017
- Confusion Matrices
  May 3, 2017
- Comparison of Different Models
  April 29, 2017
- Logistic Regressions
  April 29, 2017
- SVMs
  April 29, 2017
- Decision Trees
  April 29, 2017
- Linear Regressions
  April 29, 2017
- Getting and Processing Data
  April 29, 2017

- GitHub Integration
  April 26, 2017
- Part 1 Overview
  April 24, 2017
- Planning for the rest of the project
  April 24, 2017
- Regression (New)
  April 14, 2017
- Cluster
  April 13, 2017
- Regressions and Some Plots
  April 9, 2017
- Minimized Data
  April 8, 2017
- Correlation Matrix & Scatter Plot Matrix
  March 24, 2017
- Filtered Missing Data
  March 24, 2017
- Merged Data
  March 23, 2017
- Code for Histograms
  March 18, 2017
- Histograms
  March 18, 2017
- Data Downloaded
  March 18, 2017
- Project Proposal (v3)
  March 18, 2017
- Project Proposal (v2)
  March 2, 2017
- Project Proposal
  March 1, 2017