Write-up for Data Story Telling Project

1 Overvew

In this project, we analyze how the use of ICT (information and communication technology) at school can affect the education level, mainly on mathematics, of 15 years-old students in OECD (Organization of Ecolonic Cooperation and Development) member countries. For this purpose, we use the data of PISA 2012 available here. The dictionary for the data is given in here. The further detail of this survey and explanation of the variables in this dataset are available here.

The Tableau Stories we have created are available at:

1. **Version 1** (before the feedback is reflected)
   (To enjoy the animation effect fully, install twbx file from here.)
2. **Final Version** (after the feedback is reflected)
   (To enjoy the animation effect fully, install twbx file from here.)

2 Summary

In this visualization, we first explained that our analysis captured a tendency that the students using ICT at school frequently get low scores in math, science and reading. Our furhter analysis on math shows that the frequent use of ICT in math lessons gives a negative effect on the score of math and that the students with frequent computer simulation habit tend to get low scores in math. Based on this, one might conclude that ICT use at school is bad for math education, but the data at the same time indicates that the frequent ICT use in math lesson leads to high mathematical interest and intention of the students.

3 Design

Here is a brief explanation on the design of the visualization we have chosen. Let us explain page by page in the Tableau Story (final version):

- **(page 1)**
  To explain (1) many students start using computer at their very early ages and (2) many schools have already installed computers, we created two bar plots. I used a filter so that the audience can click a OECD member country in the map and see the bar plots for the chosen country. Based on the feedback, in order for the audience to easily see the ratio of each entry in the bar plots, we changed the x-axis to the percentage of total valid replies and added the percentage label for each entry. Typos are also corrected.

- **(page 2)**
  In this part, we raised the main question of this Tableau Story. We also added a pie chart explaining that the majority of the students are positive about the use of ICT for school work itself. Based on the feedback, we added the label for percentage of total valid replies to the pie chart and what ICT stands for.

- **(page 3)**
  In this part, we introduced three simple histograms exlaining that the students using ICT at school (for general school work purposes) frequently tend to get low scores in mathematics, science and reading. Based on the feedback, we added what the large/small index means in the plots.

- **(page 4)**
  To further pursue the correlation between the use of ICT and mathematics education, we have created two plots: (1) a histogram explaing the frequent use of ICT in mathematics lesson at school (more specific than page 3 where general ICT use for school work at school was considered) leads to low scores in mathematics. (2) To see more in detail, we have taken a specific data in PISA 2012 survary on computer simulation, and we have created a scatter plot describing that relation between the frequency of computer simulation and score in mathematics. An interactive interface is added so that the audience can click the buttons by themselves to see that the average score in math indeed becomes lower as the students have more frequent habit of doing computer simulation. Based on the feedback, we added what the large/small index means in the plot.

- **(page 5)**
  In this part, to stress that the use of ICT at school can still have positive consequence on math education, we have created two histograms showing that the frequent use of ICT in mathematics lesson at school leads to high mathematical interest and intention of the students. Based on the feedback, we also changed the color of the text "high" from blue to red and added what the large/small index means in the plots.

- **(page 6)**
  Based on the feedback, we have added the summary slide to stress the main points of our Tableau Story. Some outlooks are also added for future work.

4 Feedback

We have received the several feedbacks (from one person). Here we summaize the feedbacks themselves and how we improved our visualization based on them:

- **(page 1)**
  Typo in the text above the slide, "tudents" and "Many school has".
  $\rightarrow$ We replaced them respectively by "students" and "Many schools have".

- **(page 1)**
  It is better to add percentage of each entry to the bar plots.
  $\rightarrow$ We added the percentage labels to the bar plots (and changed labels of x-axes and tooltips accordingly).

- **(page 2)**
  It is better to add percentage of each entry to the pie chart.
  → We added the percentage labels to the pie chart (and changed the tooltip accordingly).

- **(page 2)**
  Add what ICT stands for.
  → We added a text explaining that ICT means Information and Comunication Technology.

- **(page 5)**
  Highlight the work "high" with blue color is not good (since blue intuitively implies "low").
  → We changed it to red (all the highlighted words are in red now).

- **(page 3, 4, 5)**
  When an index is used for the histograms, it is better to add texts explaining what the small/large index means.
  → We added short texts near the axes of the plots to explain the meaning of large/small index.

- It will be nice if there is a summary slide in the end.
  → We added a new slide for the summary in the end and summarized the key point of our Tableau Story as well as some outlooks.

5 Resources

For the visualization of this project, we have used the dataset of PISA 2012 survay `pisa2012.csv`. The dictionary for the variable in this file is given in a separate file `pisadict2012.csv`. Since this original datafile is huge (2.75 GB), carrying out the visualization with Tableau is slow and frustrating. We thus created the python code `extracting_pisa2012.ipynb` to extract some variables that are interesting for our purpose. After running the python code, the extracted data and corresponding dictionary are respectively stored in the newly created csv files `pisa2012_extracted.csv` and `pisadict2012_extracted.csv`.

Here is the list of the data source and some extra files we have used for the visualization:

- `pisa2012.csv`
  (**This is the original data file of PISA 2012 downloaded from UDACITY website.**)
  This csv file stores the original data of PISA 2012 survay. We note that this file is huge (2.75 GB).

- `pisadict2012.csv`
  (downloaded from UDACITY website.)
  This file stores the dictionary for the variables in the `pisa2012.csv` (`CNT` corresponds to country, etc.)

- `extracting_pisa2012.ipynb`
  This python code is used to extract from `pisa2012.csv` and `pisadict2012.csv` some interesting variables and dictionary for them for creating our visualization.

- `pisa2012_extracted.csv`
  (**This data file is used for creating the visualization in this project.**)
  This csv file stores the extracted data from `pisa2012.csv` for the purpose of our visualization.

- `pisadict2012_extracted.csv`
  This csv file stores the extracted dictionary from `pisadict2012.csv` for the purpose of our visualization.

For the further details of the survey and the variables in the dataset, we have refereed to this webpage of PISA 2012. For the detail of OECD, we have referred to Wikipedia:OECD.