Project: What Advice can be Given to Parents to Improve Students School Performance?

— A Hands-on Project on the 2015 PISA Dataset

Table of Contents

Introduction

Among the first questions a new parent encounters in the course of upbringing their child is about education. How do I teach, What do I teach, When do I teach, Do I need to teach in the first place or is the school enough? Assuming that the school is mainly responsible for education, how can I, as a parent, support the learning process will form the starting question for the following analysis.

From an economics perspective it is almost common knowledge that education is one of the main factors of national development. However, education does not only matter on a national scale, in the nearer personal scale, education is even more important some might argue. Education, on all sizes of scale, is among the main causes, of poverty, success, be that career-wise or social, or freedom, as one will have a hard time communicating when traveling, should he be equipped with a single language only.

As Education is crucial to many aspects in life, this paper will try to answer the question, what advice can be given to parents with regards to increasing their childs educational performance. The paper will derive the importance of some factors to education from the PISA dataset. PISA is short for *Programme for*

International Student Assessment and evaluates the educational levels across many countries, by testing for knowledge and skills of 15 year olds on mathematics, science and the language of the country.

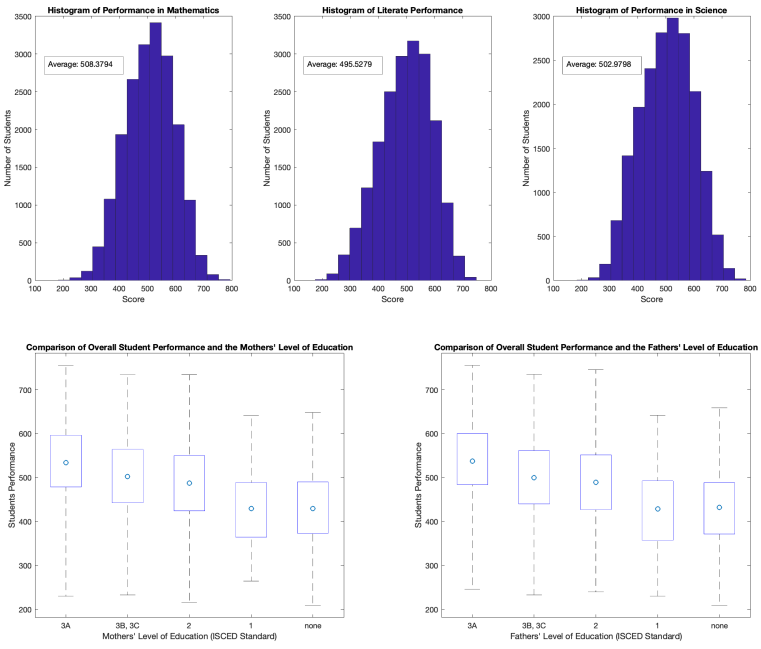
This paper is divided into four subsections, which were created based on intution and meaning, factors will be analyzed which could intuitively have an impact on educational success. Those are first, the performance of students based on their parents' level of education. Then, general inclusion of a teenager at school and how his social interactions impact his performance. This aims at giving general information on whether the subject feels like an outsider, lonely, has difficulties integrating at school or if he rather feels that he has many friends and feels like he belongs at school. The third subarea of our study targets how the relationship between parents/teacher and the students are. The goal here is to analyse if students that are encouraged more, have parents that help them, integrate in the scholar student impacts their performance at school. Fourth, we analyse how much time teenagers spend in their respective classes and more importantly in front of the internet and how that might affect their educational level (performance). Finally, we research the influence of objects in a child's home. This is, if the number of books or the presence of a video game console influence the performance of a child.

Data Description

The students that are analysed in this section are in the 8th and 9th grade, this means that they are about 13 to 15 years old in general cases. This part aims at showing the general performances of the students and preparing the data to be analyse more in details.

Remark: All of the following boxplots have been modified to display the average of the respective subgroup of data as a circle. This is quite a significant difference to usual boxplots, which plot the median instead of the mean. For our analyses the average performance however is of importance, contrary to the value of the median which represents the middle value of the dataset.

```
appendix % this is a separate file foregoing any importing and preparation steps and is
```



ISCED (International Standard Classification of Education) **Levels** (PISA 2015 Technical Report, 2017, p. 298):

1 ≙ Primary Education

2 ≙ Lower Secondary Education

3B, 3C ≙ Vocational/Pre-Vocational Upper Secondary Education

3A ≙ General Upper Secondary Education

The upper three histograms show the approximate distribution of the performances, which show that most people have a performance around 500 in the respective subjects. Some are better, some are worse and only very few perform outstandingly well or incredibly badly. The lower boxplots show that students' performance in considerably better of higher educated parents. Therfore, it would be nice to see which tools these better educated parents do which help their children perform better. As this is quite a vast undertaking, which might extend beyond the scope of this project, the analysis in Part 2 will focus on which factors generally impact students performance neglecting parents education.

Part 1: Data Preparation

Data Import

Before proper analysis can be conducted, the data needs to be imported and prepared for the analysis in such a way that it is easy to process, which means a good data type to store a large amount of numerical values with descriptive headers.

One of those is a table, which will be created according to the following steps.

1. load the raw data from PISA
2. select the columns we want to analyze
3. load selected columns into a separate table (*our_table)*

Remark: The original dataset from PISA (Student questionnaire data file — CY6_MS_CMB_STU_QQQ.sav) has been modified in JMP. All observations from Austria, Germany and Switzerland have been extracted into a separate .jmp file, which has then been exported into an Excel-Workbook.

```
warning('OFF', 'MATLAB:table:ModifiedAndSavedVarnames') % code that suppresses unwante
raw_numeric_data = readtable('CY6_MS_CMB_STU_QQQ_all_numeric.xlsx');
% load raw_numeric_data.mat; %for quicker loading, uncomment this command and
% comment out the readtable commands above.

intr_col = {'PV1MATH' 'PV10MATH' 'PV1READ' 'PV10READ' 'PV1SCIE' 'PV10SCIE' 'ST003D02T'
            'ST003D03T' 'ST034Q01TA' 'ST034Q06TA' 'ST100Q01TA' 'ST100Q05TA' 'ST123Q01NA
            'ST123Q04NA' 'ST097Q01TA' 'ST097Q05TA' 'IC006Q01TA' 'IC007Q01TA' 'ST059Q01
            'ST059Q03TA' 'ST012Q01TA' 'ST013Q01TA' 'IC001Q01TA' 'IC001Q11TA'};
% the column names are the first and last column which constitute the range of columns
% the data regarding the areas of interest (selected questions).

intr_col_ind = [];
for i = 1:length(intr_col)
    % Translating the column names into for slicing usable indices.
    y = find(strcmpi(raw_numeric_data.Properties.VariableNames,intr_col(i)));
    intr_col_ind(end+1) = y;
end
```

```
% Additional columns that have been selecte separately as they are not part of a range
add_intr_col = {'ST022Q01TA' 'ST001D01T' 'ST005Q01TA' 'ST007Q01TA'};
add_intr_col_ind = [];
for i = 1:4
    y = find(strcmpi(raw_numeric_data.Properties.VariableNames,add_intr_col(i)));
    add_intr_col_ind(end+1) = y;
end

% appending the columns
start_ind = 1:2:24;
our_table = [];
for i = 1:length(start_ind)
    temp_table = raw_numeric_data(:,intr_col_ind(start_ind(i)):intr_col_ind(start_ind(i
    our_table = [our_table temp_table]; % contains now the columns of interest
end

for i = 1:4
    add_column = raw_numeric_data(:,add_intr_col_ind(i));
    our_table = [our_table add_column];
end
```

Preparation of the Performance Data

The Center for International Student Assessment, computes multiple estimate values for the students performance, to increase accuracy of the performance assessment. There are 10 estimates per test area (test language, mathematics, science) which need to be aggregated to their means to simplify the comparison.

```
performances = our_table(:,1:30);
performances_mx = table2array(performances);
mn_math = mean(performances_mx(:,1:10),2);
mn_read = mean(performances_mx(:,11:19),2);
mn_scie = mean(performances_mx(:,20:30),2);
means = [mn_math mn_read mn_scie];
overall_performance = mean(means,2);
means(:,end+1) = overall_performance;
means_table = array2table(means);
means_table.Properties.VariableNames = {'MEAN_MATH' 'MEAN_READ' 'MEAN_SCIE' 'OVRALL_PR
```

Missing Values

Another challenge in preparing data for analysis is the confrontation with 'holes' in a dataset, meaning a missing value. As time is scarce, and computer also not always 100 percent reliable, questions in questionnaires get skipped or measurements not fully executed, resulting in an empty cell in an observations-variable table.

Multiple approaches are possible such as (a) mean substitution,(b) moving mean substitution, (c) last observation carried forward,(d) maximum likelihood and many more. In plain english those mean to (a) fill the missing data in with the sample mean, (b) fill the missing data with a mean created by a few previous data points, (c) fill the missing data with the last measured value and (d) fil the missing data with the value that is most likely, has been measured most of the time.

If all approaches fail, due to circumstances, there is one final approach to deal with missing data, being to delete the observation. But, the deletion of observation should only be done with care. In case the missing values, occur systematically, which might be due to a technical issue within a computer or occure due to the framing of a question in a way that most of the participants do not understand the question, the deletion might be critical. Should it be the case the variable, with the missing values, is highly significant and thus, relevant for the final result, the deletion might be critical as it might create a bias/distortion in the data, if the size of the data set is not large enough to counter balance.

Our dataset (our_table) however, contains the answers to very simple questions, created specifically for children of the age of 15, which is why a systematic error can not be reasonable assumed. Furthermore, as our dataset contains close to 20'000 observations, the deletion of 2-5'000 observations should not distort the data in a significant way. Any deletion of observations in the following analysis will be disclosed openly, to maintain credibility.

Remark on the Boxplots Used

The boxplots in the following parts have been altered in a significant way compared to usual boxplots.

Usual boxplots display the Median as the 'middle-value'. As the follwing analysis however, is interested in average performance and not the middle performer, the median has been sustituted by the mean, which is important to keep in mind when looking at the following plotss.

Part 2: Analysis

In the following sections the analysis in the four subgroups of interest will be performed.

Interest and Support of Superiors

Parents — the first people we ever get to know, and the ones who will have the most impact in the first few years of life and a considerable impact on further educational development. Besides the parents, there is another group of adults who will have a significant impact on a childs development, namely being the ones who take responsibly whenever the parents do not. That is the case in kindergarden, when the qualified educators take over and are the caregiver in charge, at university when professors are the ones who are the most important caregivers in the educational context, or at school, when teachers do what they are paid for, educating in various areas. Be that math, social contact, how to cook, and many more.

This section therefore will analyze the relation between the support and interest of superiors and childrens' performance.

he following questions have been selected for this section:

- The teacher shows an interest in every student's leearning.
- The teacher gives extra help when students need it.
- The teacher helps students with their learning.
- The teacher continues teaching until the students understand
- The teacher gives students an opportunity to express opinions.
- My parents are interested in my school activities.

- My parents support my educational effort.
- My parents support me when I am facing difficulties at school.
- My parents encourage me to be confident.
- Students don't listen to what the teacher says.
- There is noise and disorder.
- The teacher has to wait a long time for students to quiet down.
- Students cannot work well.
- Students don't start working for a long time after the lesson begins.

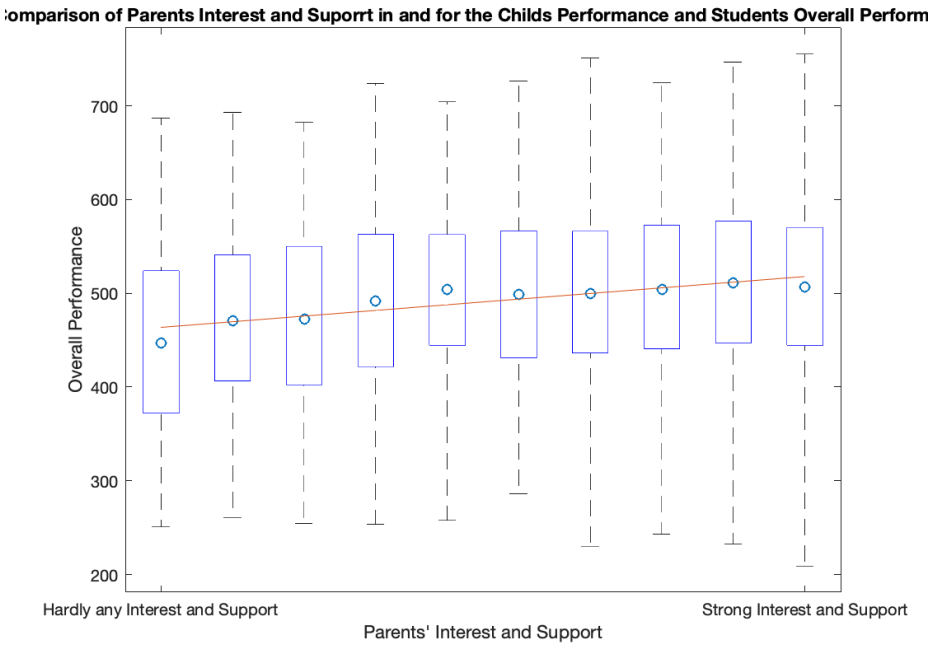
Support and Interest of Parents

Most parents are usually concerned about their childrens life decisions and their academic and future careers and therefore show interest in their childrens education and support them wherever it is feasible. The following boxplot shows the correlation between the interest and support of the parents and their childrens performance at school.

```
figure;
parent_data = [our_table(:,44:46)];
parent_data_summarized = sum(parent_data{:,1:3},2);
parent_data_refined = [array2table(parent_data_summarized, 'VariableNames', {'PARENT_I
figure;
my_boxplot(parent_data_refined)
```

```
This dataset contains 94.0478% of the original dataset,
 which constitutes to 18218 of 19371 observations.
```

```
title('Comparison of Parents Interest and Suporrt in and for the Childs Performance an
ylabel('Overall Performance')
xlabel('Parents'' Interest and Support')
xticks([1 10]);
xticklabels({'Hardly any Interest and Support' 'Strong Interest and Support'});
```



The Boxplots show the correlation between the support, interest and involvement of the parents in their children's studies and the overall performance of their children. The questionaire divides the interest, support and and encouragement into four subquestions. However, here the subquestions are combined by taking the mean of the four values, in order to simplify the process and make it easier to understand. The data shows that in the most cases, showing more interest and supporting the children also tends to result in better performances at school.

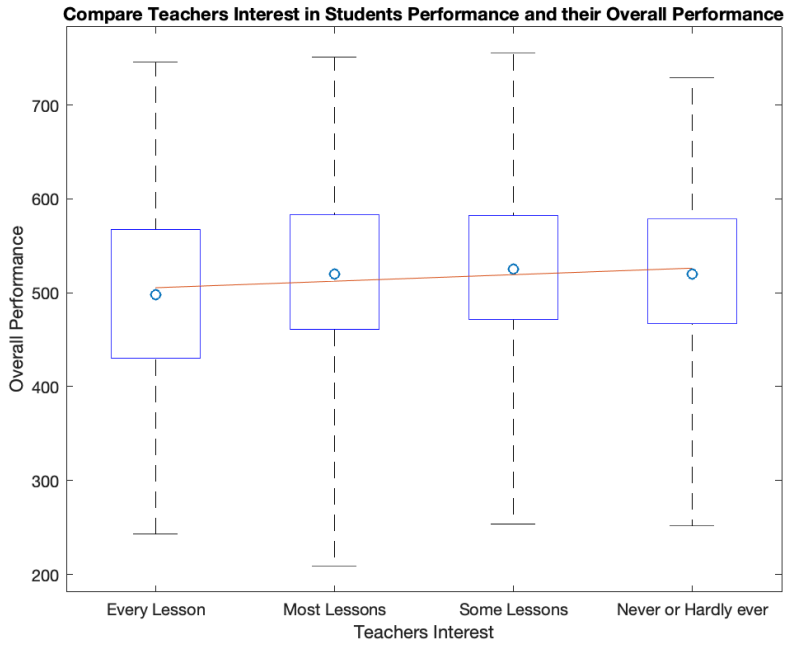
Teachers' Interest

Let us now analyse if the same holds for the attention and dedication that teachers give to students.

```
teach_int_data = [our_table(:,39) means_table(:,4)];
figure;
my_boxplot(teach_int_data)
```

```
This dataset contains 75.7731% of the original dataset,
 which constitutes to 14678 of 19371 observations.
```

```
title('Compare Teachers Interest in Students Performance and their Overall Performance
ylabel('Overall Performance')
xlabel('Teachers Interest')
xticks([1 2 3 4]);
xticklabels({'Every Lesson' 'Most Lessons' 'Some Lessons' 'Never or Hardly ever'});
```



Apparently, the teachers interest is less relevant than previously anticipated. The data shows hardly any relationship. If any, a extremely little negative relationship, which does not make sense under current assumptions that having an interested superior, boosts performance. A possible explaination might be that teachers have the task to help the whole class to reach a certain educational target at the end of the year. In order for the whole class to reach such targets, teachers sometimes have to focus more and therefore show more interest and give more support to the students who lack performance and struggle at school while the students with a better performance can work independently and receive "less support".

Teacher's Authority

The next section on the syllabus of this paper is teachers authority. The following three questions have been selected in this section as they can be interpreted as factors lying in the teachers power to control and ensure and are intuitively seeming to have a relationship with students performance.

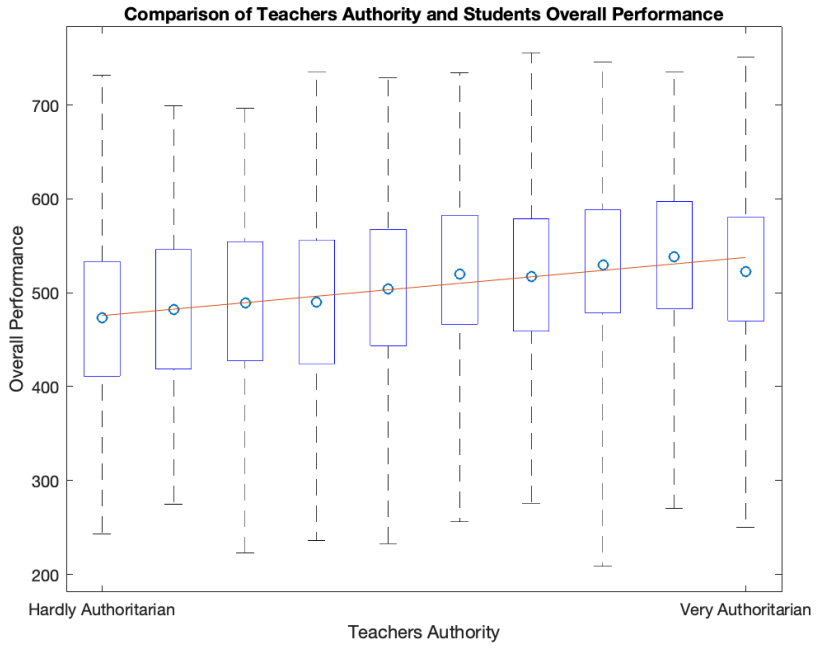
- Do students listen to what teachers say?
- Is there noise and disorder?
- Does the teacher have to wait a long time for students to quiet down

The work environment in every classroom is partly determined by the authority of the teacher and to what extent the teacher can control the class. Therefore it wil be analysed in the following section if a more authoritarian teacher benefits the performance of the students. The authority of a teacher is important as in influences the starting and ending time of lectures, the noise and disorder and might help achiving a work environment without distractions.

```
authority_data = [our_table(:,48:50)];
authority_data_summarized = sum(authority_data{:,1:3},2);
authority_data_refined = [array2table(authority_data_summarized, 'VariableNames', {'AU
figure;
my_boxplot(authority_data_refined)
```

```
This dataset contains 78.2045% of the original dataset,
 which constitutes to 15149 of 19371 observations.
```

```
title('Comparison of Teachers Authority and Students Overall Performance')
ylabel('Overall Performance')
xlabel('Teachers Authority')
xticks([1 10]);
xticklabels({'Hardly Authoritarian' 'Very Authoritarian'});
```



The data shows that the average performance increases with increasing levels of authority. It shows that a stricter lead of teachers do coincide with higher performances.

An intersting question, though beyond this papers goal, would be to look at why this is. Is it that stricter teaching, let's students learn and understand better or is it about the culture that is present when learning meaning being disciplined.

For now, the conclusion is simply that the stronger the autority the higher the performance.

Teachers Autority — deepened

Remark: Revisit this chapter after reading the chapter on noise and disorder.

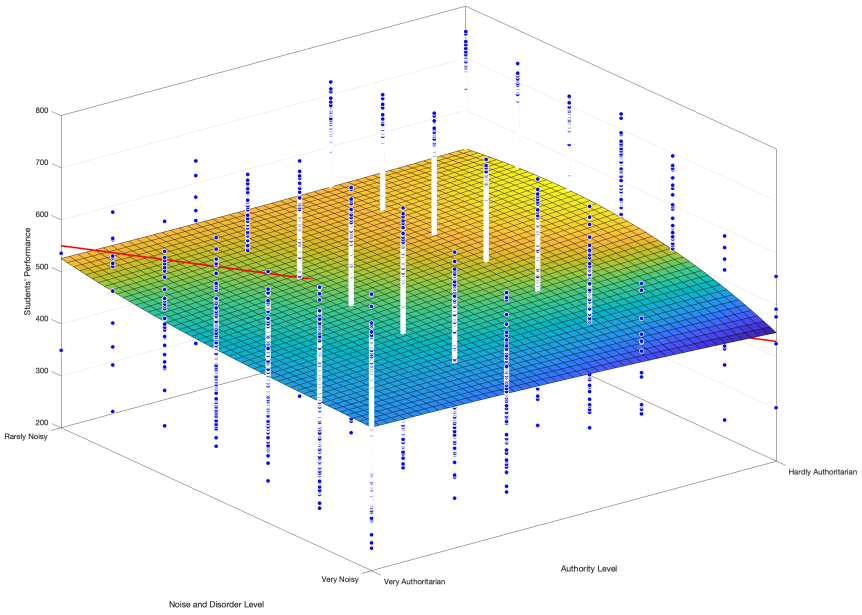
As we have seen that teachers authority had a positive relationship with students performance, remained curious however on how that could be the case precisely, the following will now try to narrow down the possible effects from higher authority ultimately affecting students performance.

```
% this section requires the curve fitting toolbox
load 'perf_nd.mat'
authority_data_2 = [our_table(:,[48,50])];
authority_data_summarized_2 = array2table(sum(authority_data_2{:,1:2},2));
deep_authority = [perf_nd(:,1) authority_data_summarized_2(:,1) means_table(:,4)];
deep_authority = clean_table(deep_authority);
```

```
This dataset contains 78.2045% of the original dataset,
 which constitutes to 15149 of 19371 observations.
```

```
X = deep_authority{:,1};
Y = deep_authority{:,2};
Z = deep_authority{:,3};
f = fit( [X, Y], Z, 'poly23' );

figure;
set(gcf, 'Position', [100, 200, 1200, 900]);
plot(f, [X,Y], Z)
hold on
xlabel('Authority Level')
xticks([1 4]);
xticklabels({'Very Authoritarian' 'Hardly Authoritarian' });
ylabel('Noise and Disorder Level')
yticks([2 8]);
yticklabels({'Very Noisy' 'Rarely Noisy'});
zlabel('Students'' Performance')
hold on
% points created to show direction of slope of plane.
hint_points_X = [4 1].';
hint_points_Y = [2 8].';
hint_points_Z= [430 550].';
plot3(hint_points_X, hint_points_Y, hint_points_Z, 'LineWidth', 2, 'Color', 'r')
hold off
```



This plot might be a little more difficult to understand, at first. Essentially, it is a graphical illustration of how the three variables, being teachers' authority, noise levels and students' performance coincide.

At all authoritarian levels, the relation between higher noise and lower performance remains. In turn when looking at all noise levels, one can see that generally higher authority coincides with higher performance, regardless of the authority level.

Another look at this plot shows, that, assuming that authority level and the noise level are perfectly correlated, resulting in a 45 degree line on the x-y plane, one can notice a non-negligible increase of the plane in this direction. This shows the combined relationship of authority and noise levels on the performance.

Social Aspects

Social interactions are said and have been proven to be as important as a good teacher and sleep (Hurst, Wallas, Nixon (2013)). Social Interaction have been shown to be useful to students in terms that they improve *"learning by enhancing their knowledge of literacy and teaching and their critical thinking and problem-solving skills"*. The goal of this section will be to analyze this phenomenon with the given data

The following questions have been selected for this section:

- I feel like an outsider (or left out of things) at school.
- I make friends easily at school.
- I feel like I belong at school.
- I feel awkward and out of place in my school.
- Other students seem to like me.
- I feel lonely at school.

Comfortableness at School

Famous psychologist Maslow proposed a fundamental theory on the hierarchy of needs. In his theory before proper social contact can be made, so he says, physiological and safety needs need to be satisfied first. Physiological needs are here assumed to be satisfied, which brings us to the safety requirement. Admittedly,

safety per se is quite a reach. Similar to the need of feeling safe when one leaves his house, this is applied to ones need of feeling 'safe' and comfortable in their school environment, which will be analyzed in the following.

```
social_data = [our_table(:,33:38) means_table(:,4)];

belonging_data = [social_data(:,3) means_table(:,4)];
awkwardness_data = [social_data(:,4) means_table(:,4)];
loneliness_data = [social_data(:,6) means_table(:,4)];

figure;
set(gcf, 'Position', [100, 200, 1200, 420]);

subplot(1,3,1)
my_boxplot(belonging_data)
```

```
This dataset contains 93.5109% of the original dataset,
 which constitutes to 18114 of 19371 observations.
```

```
title({'Comparison of Students Feeling of Belonging';'to School and their Overall Perfo
xlabel('Belong at School')
xticklabels({'Yes' 'I think' 'I do not know' 'No' })
ylabel('Overall Students Performance')

subplot(1,3,2)
my_boxplot(awkwardness_data)
```

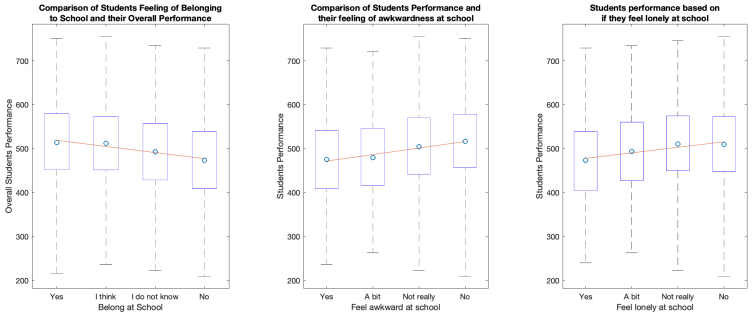
```
This dataset contains 93.1289% of the original dataset,
 which constitutes to 18040 of 19371 observations.
```

```
title({'Comparison of Students Performance and'; 'their feeling of awkwardness at schoo
xlabel('Feel awkward at school')
xticklabels({'Yes' 'A bit' 'Not really' 'No' })
ylabel('Students Performance')

subplot(1,3,3)
my_boxplot(loneliness_data)
```

```
This dataset contains 93.7432% of the original dataset,
 which constitutes to 18159 of 19371 observations.
```

```
title({'Students performance based on'; 'if they feel lonely at school'})
xlabel('Feel lonely at school')
xticklabels({'Yes' 'A bit' 'Not really' 'No' })
ylabel('Students Performance')
```



The data proves this connection, as all three questions show a difference of approx. 40 points between those who are comfortable at school and those who are not, which intuitively does make sense as well. Therefore, a students state of mind at school and integration seem to paly an important role in a childs perfromance at school.

Social Connections

Also in connection with Maslow's hierarchy of needs, namely the belongingness and love needs and maybe esteem needs, the next group of aspects, seeming to be of relevance intuitively, are aspects concerning social connections such as, friends and likedness, which will be analyzed in the following.

```
outsider_perf = [social_data(:,1) social_data(:,end)];
friends_data = [social_data(:,2) means_table(:,4)];
liked_data = [social_data(:,5) means_table(:,4)];

figure;
set(gcf, 'Position', [100, 200, 1200, 420]);

subplot(1,3,1)
my_boxplot(outsider_perf);
```

```
This dataset contains 94.1665% of the original dataset,
 which constitutes to 18241 of 19371 observations.
```

```
title('Comparison of Student Performance and Outsider Perception');
xlabel('Feeling Like an Outsider');
xticklabels({'Yes' 'Yes a little' 'Not really' 'No' });
ylabel('Students Performance');

subplot(1,3,2)
my_boxplot(friends_data)
```

```
This dataset contains 93.9394% of the original dataset,
 which constitutes to 18197 of 19371 observations.
```

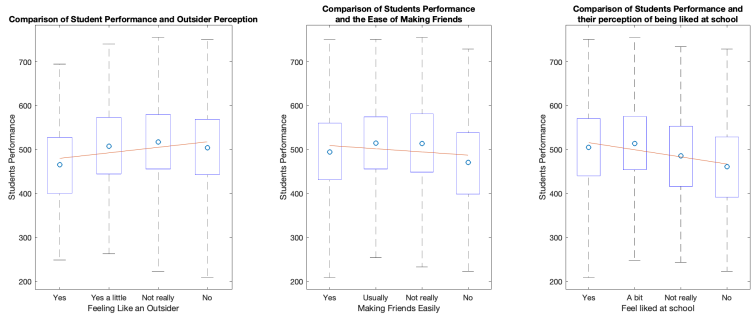
```
title({'Comparison of Students Performance'; 'and the Ease of Making Friends'})
xlabel('Making Friends Easily')
xticklabels({'Yes' 'Usually' 'Not really' 'No' })
ylabel('Students Performance')

subplot(1,3,3)
```

```
my_boxplot(liked_data)
```

```
This dataset contains 93.4232% of the original dataset,
 which constitutes to 18097 of 19371 observations.
```

```
title({'Comparison of Students Performance and'; 'their perception of being liked at s
xlabel('Feel liked at school')
xticklabels({'Yes' 'A bit' 'Not really' 'No' })
ylabel('Students Performance')
```



Again, social aspects appear not to be negligible in our performance driven world.

Feeling like an outsider, making friends not that easily, which coincides with a less social appearance, be that visibly or personality-wise, or not feeling liked at school all have a non-negligible connection with performance in school. Reason for that might be, that if one feels relaxed and comfortable at school, one can spend full effort in learning, whereas someone who does not, might have to fumble with the uncomfortable feelings which hinders their learning process and ultimately their performance. However further research and data would be needed for deeper analysis.

Exposure to Potential Factors

This section focuses on anaylsising how much time a student spends on different activities and how this influences a childs grades. The main quetions looked at are the following:

- During a typical weekday, for how long do you use the Internet outside of school?
- On a typical weekend day, for how long do you use the Internet outside of school?
- How many class periods per week are you typically required to attend in (a) your test-language (b) mathematics and (c) science?
- What language do you speak at home most of the time?
- What is the highest level of schooling completed by your mother (father)?

Time Spent on the internet

With all the possibilities on the internet, from watching videos, playing games with your friends on facebook, to browsing and shopping, there are a lot of time-consuming options which, one might think, would have little positive effects on grades. With an outlook on an ever increasing level of digitalization in various societies, this could be reason for worries on the future, as students might increasingly spend time on useless internet games instead of learning to solve linear algebra — or could it?

This section aims to shed some light on this contemporary issue.

```
inet_week = [our_table(:,53:54) means_table(:,4)];
```

```
inet_week.Properties.VariableNames = {'WEEKDAY' 'WEEKENDDAY' 'OVRALL_PRFMNC'};
inet_week = clean_table(inet_week);
```

```
This dataset contains 62.6504% of the original dataset,
 which constitutes to 12136 of 19371 observations.
```

In order to be able to compare internet use over the whole week and the performance, the separate values for internet use on weekday and weekenddays need to be concatenated.

Now that we have unique values for the combined total duration people have spent on the internet over the course of a whole week, we need to assign those values the specific duration.

Therefore, we first calculate all possible values of total internet usage across a week.

```
whole_week = 5*(10.^inet_week.WEEKDAY) + 2*(3.^inet_week.WEEKENDDAY); % different bases
tot_duration_value=[];
a = [0 0.25 0.75 1.5 3 5 6]; % the possible durations on weekdays
b = [0 0.25 0.75 1.5 3 5 6]; % the possible durations on weekenddays
% goal of loop: calculate all combinations of total durations.
for i = 1:7 % first for loop to loop through all weekdays
    z = a(i);
    for l = 1:7 % second for loop to loop through all weekenddays
        z2 = b(l);
        tot_duration_value(end+1) = z+z2; % total duration calculated
    end
end

poss_values = unique(whole_week); % side-effect of unique() the values get sorted from
poss_durations = array2table([poss_values tot_duration_value']);
poss_durations.Properties.VariableNames = {'DURATION_VALUE' 'DURATION_IN_HOURS'};
whole_week_duration_hours = [];
for i = 1:length(whole_week)
whole_week_duration_hours(end+1) = poss_durations.DURATION_IN_HOURS(find(poss_durations
end

inet_whole_week_data = [array2table(whole_week_duration_hours.') inet_week(:,3)];
inet_whole_week_data.Properties.VariableNames{1} = 'DURATION_HOURS';

inet_in_week_data = inet_week(:,[1,end]);
inet_weekend_data = inet_week(:,[2,end]);

figure;
set(gcf, 'Position', [100, 200, 1200, 420]);
sgtitle('Comparison of Overall Performance and Time Spent on the Internet', 'fontweight

subplot(1,3,1)
my_boxplot(inet_in_week_data)
```

```
This dataset contains 100% of the original dataset,
 which constitutes to 12136 of 12136 observations.
```

```
xticks([1 2 3 4 5 6 7]);
xticklabels({'0' '0-0.5' '0.5-1' '1-2' '2-4' '4-6' '6<'});
```

```
xlabel('Hours Spent on the Internet per Weekday')
ylabel('Overall Students Performance')

subplot(1,3,2)
my_boxplot(inet_weekend_data)
```

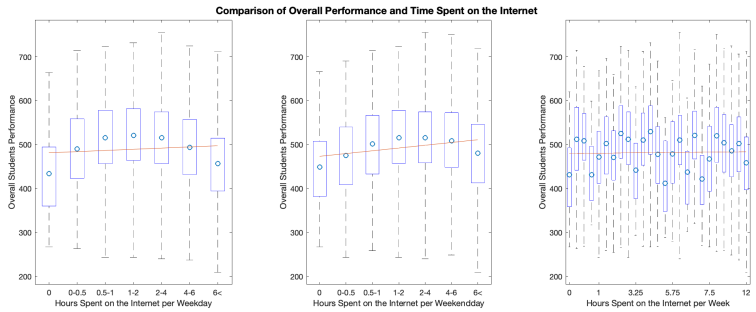
```
This dataset contains 100% of the original dataset,
 which constitutes to 12136 of 12136 observations.
```

```
xticks([1 2 3 4 5 6 7]);
xticklabels({'0' '0-0.5' '0.5-1' '1-2' '2-4' '4-6' '6<'});
xlabel('Hours Spent on the Internet per Weekendday')
ylabel('Overall Students Performance')

subplot(1,3,3)
my_boxplot(inet_whole_week_data)
```

```
This dataset contains 100% of the original dataset,
 which constitutes to 12136 of 12136 observations.
```

```
xticks([1 5 10 15 20 25]);
xticklabels({'0' '1' '3.25' '5.75' '7.5' '12'});
xlabel('Hours Spent on the Internet per Week')
ylabel('Overall Students Performance')
```



Surprisingly, these graphs shows that the relationship between hours spent on the internet per week and the students overall performance is hardly existent. Therefore, a conclusive advice on total usage of the internet cannot be derived yet.

Time Spent on Subjects

"ten thousand hours is the magic number of greatness"

— Malcolm Gladwell

This acclaimed quote from Malcolm Gladwell, a well-known author and journalist, suggests that the more hours you spent on a subject the better you get.

The further analysis aims to analyze this relationship between time spent in a subject and the performance in the respective subject.

```
math_class_time = [our_table(:,56), means_table(:,1)];
reading_class_time = [our_table(:,55) means_table(:,2)];
science_class_time = [our_table(:,57) means_table(:,3)];

figure;
set(gcf, 'Position', [100, 200, 1200, 420]);
sgtitle('Comparison of Hours Spent in a Subject and the Performance in the respective s
subplot(1,3,1)
my_boxplot(math_class_time)
```

```
This dataset contains 90.5116% of the original dataset,
 which constitutes to 17533 of 19371 observations.
```

```
xlabel('Number of Class Periods spent in Mathematics')
ylabel('Students Performance')
xticks([0 10 18 23 26]);
xticklabels({'0' '10' '20' '30' '40'});

subplot(1,3,2)
my_boxplot(reading_class_time)
```

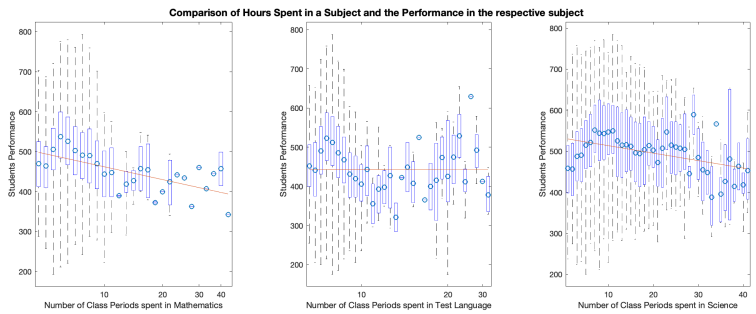
```
This dataset contains 90.6458% of the original dataset,
 which constitutes to 17559 of 19371 observations.
```

```
xlabel('Number of Class Periods spent in Test Language')
ylabel('Students Performance')
xticks([0 10 25 31 40]);
xticklabels({'0' '10' '20' '30' '40'});

subplot(1,3,3)
my_boxplot(science_class_time)
```

```
This dataset contains 89.5875% of the original dataset,
 which constitutes to 17354 of 19371 observations.
```

```
xlabel('Number of Class Periods spent in Science')
ylabel('Students Performance')
xticks([0 10 20 30 40]);
xticklabels({'0' '10' '20' '30' '40'});
```



Interestingly and contrary to intuition, one cannot conclude that the common saying "the more, the better" is correct in this context. Possibly, even the contrary is true in the range from 4 to 10 hours per week. Seemingly, from any hour more than 4 per week reduces the performance in mathematics. It is questionnable whether more hours directly reduce the performance, however, one could argue that the result is possibly due to reduced interest in the subject of mathematics or simply that due to the more indeepth and intense learning it becomes tougher to grasp all of it for students. This however, is only a hypotheses which is not proven to this point. Further research needs to be conducted on time spent on a subject and the interest to derive such a conclusion.

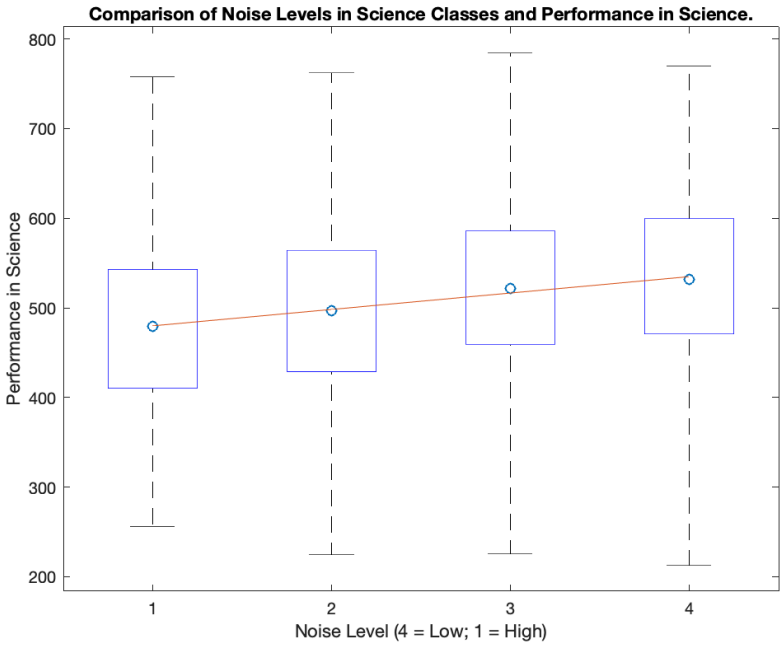
Noise and Disorder

Noise being an essential factor to a students focus and concentration, we want to depict this factor further more and analyze the exposure to noise and disorder in a classroom and its relationship with performance.

```
perf_nd = [our_table(:,49) means_table(:,3)];
figure;
my_boxplot(perf_nd)
```

```
This dataset contains 78.9324% of the original dataset,
 which constitutes to 15290 of 19371 observations.
```

```
title('Comparison of Noise Levels in Science Classes and Performance in Science.')
ylabel('Performance in Science')
xlabel('Noise Level (4 = Low; 1 = High)')
```



We see a negative relationship of noise levels and performance, which, intuitively, does make sense. Should one be sitting in a class where noise is high, how is one supposed to be able to focus among the other noisy students. (Add disorder fact to negative performance connection.)

The difference between the average performance in a very noisy class (noise and disorder in every or most lessons) and the average performance in a rarely noisy class (some/hardly any noisy classes) is roughly 40 points which is not to be neglected.

(For a deepend on noise and disorder authoritiarism and performance return to this chapter.)

Objects Present and Used at Home

Another set of factors that might contribute are the objects at a students disposal, which is what the following sections are going to analyze, questions such as;

- How many books are there in your home?
- Are any of theses devices for you to use at home (do you use them)?

-Portable laptop, or notebook

-Tablet computer (e.g. <iPad®>, <BlackBerry® Playbooks>)

-Video games console>, e.g. <Sony® PlayStation®>

-eBook reader, e.g. <Amazon® KindleTM>

Books and Student Performances

Almost a pseudo-fact in today's society is who reads is smart. In this chapter the target is to find out whether this is true, as the amount of books in the household are compared to the performances in different subjects.

```
books_data_mathperf = [our_table(:,66) means_table(:,1)];
books_data_litperf = [our_table(:,66) means_table(:,2)];
books_data_scieperf = [our_table(:,66) means_table(:,3)];
books_data_ovrperf = [our_table(:,66) means_table(:,4)];

figure;
set(gcf, 'Position', [100, 200, 1200, 420]);

subplot(2,2,1)
my_boxplot(books_data_ovrperf)
```

```
This dataset contains 95.1216% of the original dataset,
 which constitutes to 18426 of 19371 observations.
```

```
title('Comparison of Students Overall Performance and Number of Books at Home.')
xlabel('Number of Books at home')
xticklabels({'0-10' '11-25' '26-100' '101-200' '201-500' '>500'})
ylabel('Students Performance')
subplot(2,2,2)
my_boxplot(books_data_scieperf)
```

```
This dataset contains 95.1216% of the original dataset,
 which constitutes to 18426 of 19371 observations.
```

```
title('Comparison of Students Scientical Performance and Number of Books at Home.')
xlabel('Number of Books at home')
```

```
xticklabels({'0-10' '11-25' '26-100' '101-200' '201-500' '>500'})
ylabel('Students Performance')

subplot(2,2,3)
my_boxplot(books_data_mathperf)
```

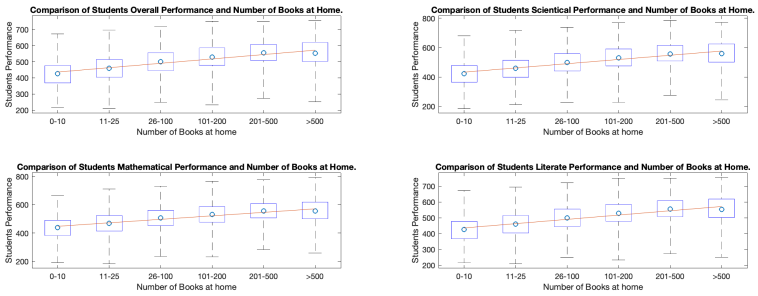
```
This dataset contains 95.1216% of the original dataset,
 which constitutes to 18426 of 19371 observations.
```

```
title('Comparison of Students Mathematical Performance and Number of Books at Home.')
xlabel('Number of Books at home')
xticklabels({'0-10' '11-25' '26-100' '101-200' '201-500' '>500'})
ylabel('Students Performance')

subplot(2,2,4)
my_boxplot(books_data_ovrperf)
```

```
This dataset contains 95.1216% of the original dataset,
 which constitutes to 18426 of 19371 observations.
```

```
title('Comparison of Students Literate Performance and Number of Books at Home.')
xlabel('Number of Books at home')
xticklabels({'0-10' '11-25' '26-100' '101-200' '201-500' '>500'})
ylabel('Students Performance')
```



Intutitively, the data proves the intuition of a relationship between higher overall performance and the number of books at home. Interestingly however, this relationship exists across all of the three disciplines, Mathematics, Science, and Literacy and not on Literacy only, as one might reasonably have assumed.

One could try to formulate the reasoning behind that multiple points. Here are a few attempts.

One, and also the simplest possibility could be, that the households do have books on Mathematics, Science and usual novels, which would mean that an average book collection has books on a variety on topics, which in turn help children, who do have access to them in their home, with their learning progress.

However, in our specialised world it might be debatable whether that is true, namely that the average houshold has books on all three of those subjects. Furthermore, should parents have books on certain subjects, it may remain doubtful to assume that those exact books, are useful on a high-school-level.

Therefore, in a more refined reasoning-attempt the increased performance across multiple disciplines does not originate from the books themselves, but rather of a 'culture' of books and reading. This 'culture' is a rather broad term, but tries to summarize the contact with knowledge itself. Given that a parent has many books, and assuming they do not exist for decoration purposes only, it could be possible that those parents

present the contact with books and knowledge in a totally different way than households with fewer, let alone no books. Unfortunately the current dataset does not allow for such analysis and needs to be forwarded.

Possession of Digital Devices

"The Attention Economy to the Addiction Economy"

— Heather West (2018)

Heather West suggests that todays attention economy, an economy evolving around stealing a person's attention as much as possible, steers towards having addictive features, which might be bad for any user. As education is somewhat related to attention, as one needs to pay attention to learn, we want to look at how digital attention-stealers, might have a relationship with student performance.

Therfore, the following questions have been selected:

```
devices_data = [our_table(:,[67,69,71,77]) means_table(:,4)];
desktop_data = [devices_data(:,[1,5])];
tablet_data = [devices_data(:,[2,5])];
gaming_data = [devices_data(:,[3,5])];
ereader_data = [devices_data(:,4) means_table(:,2)];

figure;
set(gcf, 'Position', [100, 200, 1200, 420]);
subplot(2,2,1)
my_boxplot(desktop_data);
```

```
This dataset contains 63.9255% of the original dataset,
 which constitutes to 12383 of 19371 observations.
```

```
title('Comparison of the Possestion of a Desktop and Overall Perforformance');
% xlabel('Possession/Use of a Desktop');
xticklabels({'Yes and I use it' 'Yes, but I do not use it' 'No'});
ylabel('Overall Students Performance');

subplot(2,2,2)
my_boxplot(tablet_data);
```

```
This dataset contains 63.3111% of the original dataset,
 which constitutes to 12264 of 19371 observations.
```

```
title('Comparison of the Possesion of a Tablet and Overall Performance');
% xlabel('Possession/use of tablet at home');
xticklabels({'Yes and I use it' 'Yes, but I do not use it' 'No'});
ylabel('Overall Students Performance');

subplot(2,2,3)
my_boxplot(gaming_data);
```

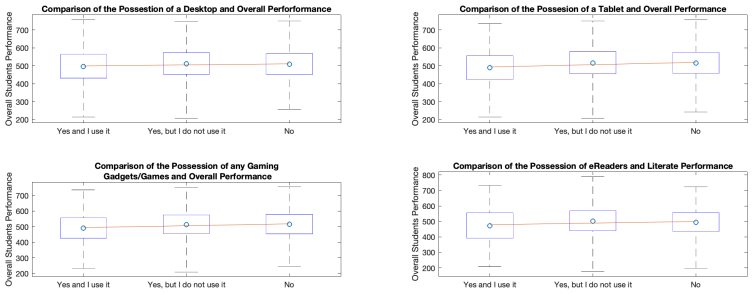
```
This dataset contains 63.4867% of the original dataset,
 which constitutes to 12298 of 19371 observations.
```

```
title({'Comparison of the Possession of any Gaming';' Gadgets/Games and Overall Perform
% xlabel('Use/possession of video games at home');
xticklabels({'Yes and I use it' 'Yes, but I do not use it' 'No'});
ylabel('Overall Students Performance');

subplot(2,2,4)
my_boxplot(ereader_data)
```

```
This dataset contains 62.8362% of the original dataset,
 which constitutes to 12172 of 19371 observations.
```

```
title('Comparison of the Possession of eReaders and Literate Performance') %changed to
% xlabel('Possesssion/Use of an ebook at home or not')
xticklabels({'Yes and I use it' 'Yes, but I do not use it' 'No'});
ylabel('Overall Students Performance');
```



Across all comparison a difference in averages can be observed. Those who do not possess the item of interest, tend to perform ≈ 20 points more.

Up to now, we extracted the effects from individual gadgets on performance. The next step shall aggregate the results across multiple gadgets.

Based on the above we notice that students who own a laptop and use it does not really have an influence on their grades. It would more likely be the numbers of hours they spend on it for leisure that would impact their performance. Then students who use tablets tend to have slightly lower grades the same holds fror students playing viedo games. Surprisingly, students that use ebooks also tend to have lower grades.

Therefore, following our analysis, laptop use does not have a major impact on students grades, however the other electronic devices tend to have a negative impact on students performances. Advice for parents is therefore to limit the number of electronic devices that their children use at home.

Total Factor Analysis

Until now we have analyzed and depicted the factors and their impact mostly individually.

What shall be attempted now is to create a summarizing model, which takes multiple factors into account. Certain Factors have been selected for their strong correlation such as:

- whatever we select.

A model which combines the individual effects into a bigger one is a so-called linear regression model. What we did before was a linear regression model as well, with one variable however, being the single factor. The following will create a so-called multivariate linear regression model, which is nothing more than what we have done before, with simply multiple factors taken into account, following the following equation:

$$y = \beta_0 + \sum_{i=1}^{n} \beta_i * x_i$$

The beta values will indicate (a) the magnitude and (b) the direction of the impact of a factor.

```
grouped = sum(table2array(our_table(:,[35,36,38])),2);
```

In order to conduct the linear regression in a proper manner, positively influencing values of variables need to be bigger than less or not affecting variables. Therfore, the numbering of some questions needs to be changed, which will be done in the following.

```
mvregress_mx = [ones(19371,1) our_table{:,[39, 49, 55,56,57,66,68,67,69,71,77]} groupec
%
size_mx = size(mvregress_mx); %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%% continue here. the ones gotta be switched to 3 and vice versa...

aux_2 = NaN(size_mx(1),1);
for i = 1:4
    aux = mvregress_mx(:,2);
    aux_2(aux == (5-i)) = i;
end
mvregress_mx(:,2) = aux_2 ;


for h = [9 10 11 12]
    for k = 1:size_mx(1)
      mvregress_mx(k,h)= (4- mvregress_mx(k,h));
    end
end

mvregress_mx_cl = table2array(clean_table(array2table(mvregress_mx)));
```

```
This dataset contains 46.3528% of the original dataset,
 which constitutes to 8979 of 19371 observations.
```

```
mvregress_vari = mvregress_mx_cl(:,[1:end-1]);
```

Remark: The methodology could be executed without the data cleaning, however would hide the representativity and in turn the authenticity, which is why it is disclosed here.

```
betas = mvregress(mvregress_vari, mvregress_mx_cl(:,end));
display(betas);
```

```
betas = 13×1
```

```
  427.6797
   -5.6340
   10.9362
   -2.1435
    1.8886
    0.3150
   25.5816
   -4.3419
   -4.2886
   -8.2083
      ⋮
```

```
deviations = [];

for i = 1:8979
    % temp_observ = array2table(mvregress_mx_cl(i,:));
    temp_dev = mvregress_vari(i,:)*betas - mvregress_mx_cl(i,end);
    deviations(end+1) = temp_dev;
end
mean_error = mean(deviations)
```

```
mean_error = 5.1722e-13
```

```
Questions = {'Constant' 'The teacher shows an interest in every student''s learning' .
    'There is no noise and disorder' 'Number of class periods per week in the test lan
    'Number of class periods per week in mathematics' 'Number of class periods per wee
    'Number of books at home' 'Having Desktop Computer' 'Having a Notebook' 'Having a
    'Having an eReader' 'Comfort at School'};
for i=1:length(Questions)
disp([num2cell(betas(i)) Questions(i)]);
end
```

```
    [427.6797]    'Constant'
    [-5.6340]    'The teacher shows an interest in every student's learning'
    [10.9362]    'There is no noise and disorder'
    [-2.1435]    'Number of class periods per week in the test language'
    [1.8886]    'Number of class periods per week in mathematics'
    [0.3150]    'Number of class periods per week in science'
    [25.5816]    'Number of books at home'
    [-4.3419]    'Having Desktop Computer'
    [-4.2886]    'Having a Notebook'
    [-8.2083]    'Having a Tablet'
    [-1.0300]    'Having a Console'
    [-6.6949]    'Having an eReader'
    [3.5502]    'Comfort at School'
```

When comparing the results received here, we see that the overall effects are no longer as strong as they were in the single comparison case. Books at home, and noise and disorder have maintained their importance in this multivariate setting, the others however, are rather suprising.

Conclusion

Connecting to the analysis above, this section aims at giving advice to parents on how they can help their child improve their performances at school.

Remark: The following advice will be drawn from correlations and assuming a causal connection.

We started by analysing the performance of students according to the amount of interest and support they received from their parents and teachers. It was observed that parents interest plays a greater role than teachers' interest in students' performances. However, teacher's authority in the class room was very relevant to our analysis. It demonstrated that a more respectful classroom (i.e. no noise, no interrupting the teacher or arriving late to class) towards the teacher tends to perform better. A sidenote; the result might be inaccurate to a certain degree, as the students have different teachers with differnt authority levels in the different subjects which are summarized here into a single variable. Despite it, the first piece of advice one can derive from the previous analysis is that parents should be involved and supportive in their children school life. Having a supportive environment to grow in seems to play a big role in a child's development and in the grades he/she can achieve.

The second part of our project focuses on the social aspects of student's life at school. More precisely, on how their state of mind regarding school in general (i.e. feeling like an outsider, making friends, belonging at school, feeling awkward at school…) influences their grades. Overall one could notice that students that have a rather negative feeling or state of mind about school and the friends they have at school are less prone to have an equally successful performance as students who feel rather good at school. Therefore, it can be advised to parent to care seriously on the attitude that their child has towards going to school. If their child does not seem to integrate properly or feels uncomfortable at school it is recommended that parents take action. That could be to talk to the school and try to integrate the child better or simply change the child's schools.

The penultimate part of our analysis dealt with the time students spend on the internet and the amount of time they spend studying different subjects. Regarding internet usage and its influence on students' grades, one could only derive a very weak correlation between both. It would seem like hours spent on the internet is not a decisive factor for scholars' grades. Similarly, based on the examination of the data more school hours of one subject does not rime with better performances. On can imagine that can be due to the increase of level in the subject making it more complicated and more difficult for students to achieve better grades or even that too many hours of one subject leads student to dislike it. In this light, no specific advice can be given to parents regarding these two topics.

Finally, this paper focused on the items that students own at home and their usage of those. One questions if the use repeated use of some of these items can have an influence on students' performances at school. Concentrating on the number of books owned at home it could be deduced that students who own more books at home seem to have larger overall performances across subjects. Based on our analysis laptops has very little influence on students' grades (one can argue that the amount spend on the laptop could), but that other electronic devices such as tablets, eBooks or gaming console has a negative correlation with students' grades. Accordingly, parents should not promote the excessive use of electronic devices with the expectation of a laptop which based on the analysis might nowadays be considered as a central piece for efficient work and can in fact participate in a student's performances at school.

In further research one should give a closer look to more advanced statistical methods, to determine releationships in the aggregate of the data and not single-variable cases only as that is, like we have seen maybe not the most exact in the real setting of multiple factors affecting performance rather than a single one. Furthermore could one try to detect inferencial connections rather than correlations, as this is neglecting the causation part of reality.

List of References

Gladwell, Malcolm, 1963-. (2008). Outliers : the story of success. New York :Little, Brown and Co.

Hurst, B., Wallace, R., & Nixon, S. B. (2013). The Impact of Social Interaction on Student Learning.Reading Horizons, 52(4). Retrieved fromhttps://scholarworks.wmich.edu/reading_horizons/vol52/iss4/5

MathWorks. (2017, July 7). Mathworks. Retrieved from How do I suppress the table warning in R2017a?: https://ch.mathworks.com/matlabcentral/answers/347777-how-do-i-suppress-the-table-warning-in-r2017a

MathWorks. (n.d.). Mathworks Documentation. Retrieved from mvregress: https://ch.mathworks.com/help/stats/mvregress.html#btkqm0f-Y

PISA. (2017). PISA Technical Report 2015. OECD.

West, H. (2018, July 27). mozilla.org. Retrieved from The Attention Economy to the Addiction Economy: https://blog.mozilla.org/internetcitizen/2018/07/27/attention-addiction/

Author’s declaration of authorship

We hereby certify that –

We have written the program ourselves except for clearly marked pieces of code –

We have tested the program and it ran without crashing (if applicable)