# PAIN POINTS OF A STUDENT

Context in portugal with an attempt to make a Singapore spin on the findings

Overview

- Problem statement:
  - To find out how we can improve student's pain points to improve overall education experience
    - Why?
    - To assist in School/government policies to improve overall education quality and education variety

- Datasets:
  - Student-mat.csv (for Math)   395 rows × 33 columns
  - Student-por.csv (for Portugese) 649 rows × 33 columns
    - Merge by columns:
      "school","sex","age","address","famsize","Pstatus","Medu","Fedu","Mjob","Fjob","reason","nursery","internet"))
      - Able to identify 370 identical participants across the 2 surveys

  - Why this dataset?
    - 3 grades for secondary school is not far off from SG's 4 grades (sec1-4)
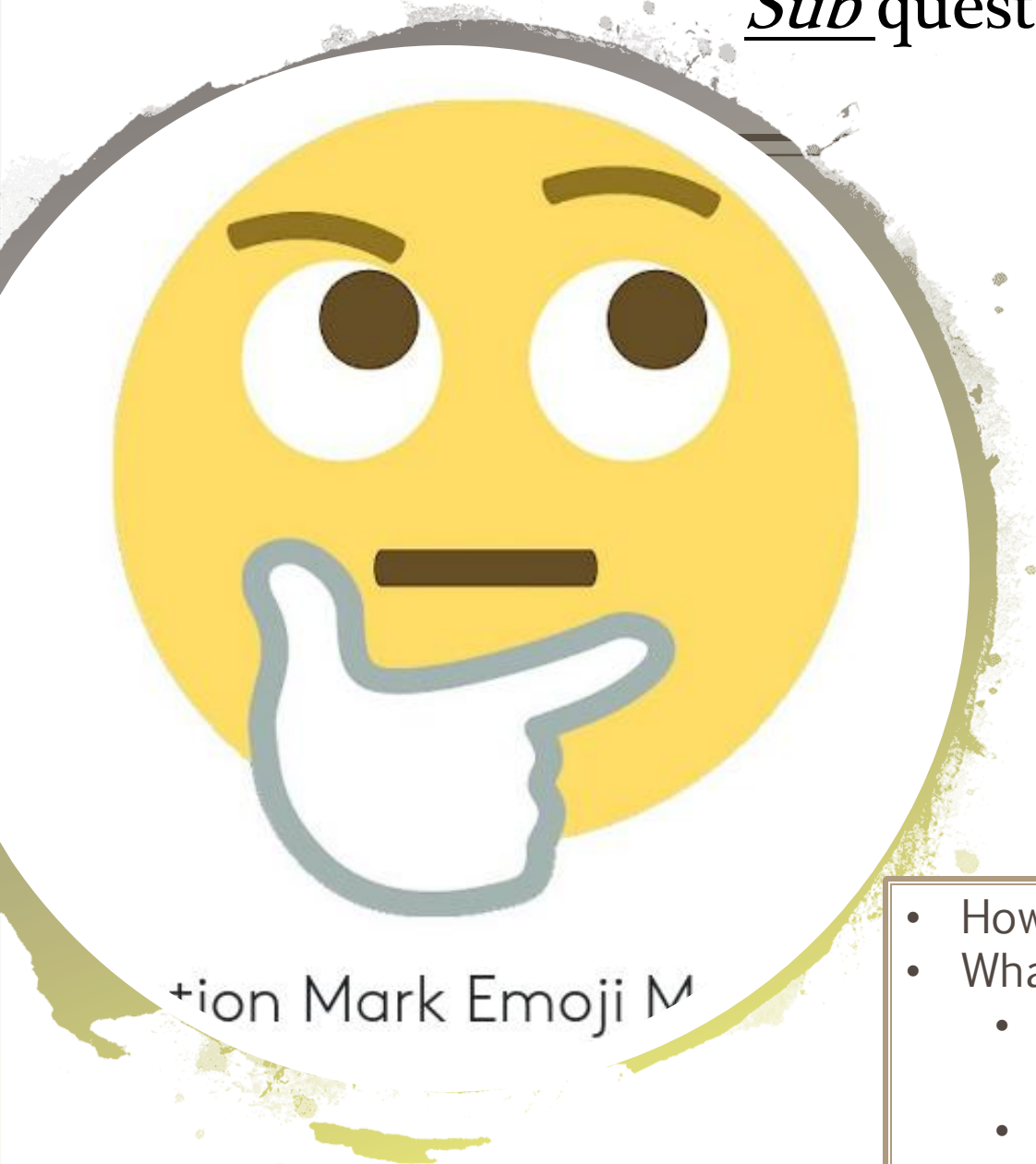
Sub questions:

- **Ideals**
  - Why either School of choice,
  - CCA/ECA options,
  - want to pursue higher education
- **Health**
  - Absenteeism,
  - Alcholism (mental health?)
- **Study/life balance**
  - Romantic relationships,
  - time spent out of school
- **internal/External support**
  - Paid tuition, school support
- **Family Background**
  - Parent's education as a proxy for income
  - are parents living together?

Main questions:

- How does it improve grades?
- What suggestions can we provide to the client?
  - Eg. Insights on pushing Singapore to a certain direction like being a Data science hub
  - Insights on whether the current govt. spending on education is sufficient and if it actually helps the more needy students

+ion Mark Emoji M

Sub-question based on home location

- Does living in Urban or rural affect grades?

```python
1  # T-test to check if GP is the more urban school
2  Ho = "Students who live in urban areas do better than R interms of passes"   # Stating the Null Hypothesis
3  Ha = "Students who live in urban areas do not do better interms of passes"   # Stating the Alternate Hypothesis
4
5  x1 = np.array(student_both.G3_x[student_both.address =="U"][student_both.G3_x > 10])
6  # Selecting students who live in Urban and went to GP
7  y1 = np.array(student_both.G3_x[student_both.address == "R"][student_both.G3_x > 10])
8  # Selecting students who live in Urban did NOT go to GP
9
10 t, p_value  = stats.ttest_ind(x1,y1, axis = 0)  #Performing an Independent t-test
11
12 if p_value < 0.05:  # Setting our significance level at 5%
```

```
Students who live in urban areas do better than R interms of passes as the p_value (0.3168604247165162) > 0.05,
 Alt hypothesis is not statistically significant and indicates,
 strong evidence for the null hypothesis as there is more than 95% probability the null is correct.
```

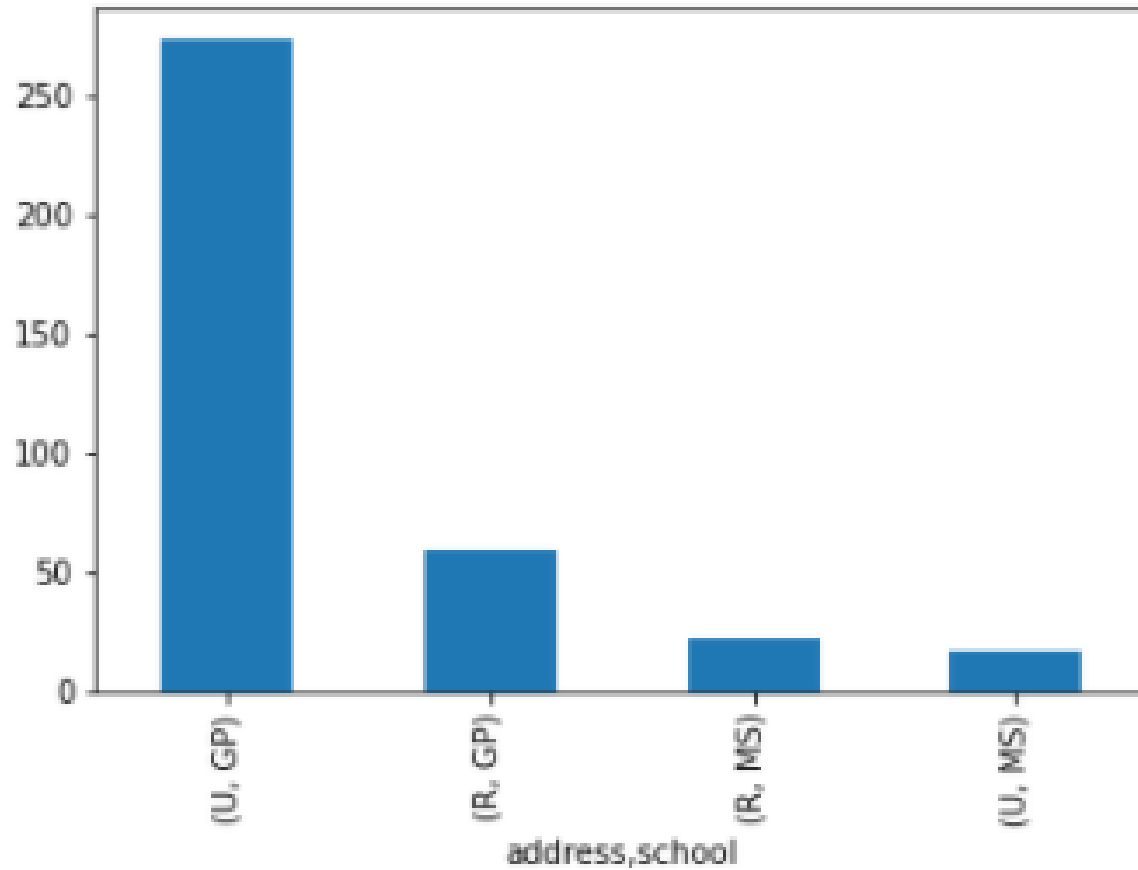# Mouzinho da Silveira High School (more rural)
(a look at the 2 schools)

Mouzinho da Silveira High School

Agrupamento de Escolas Gabriel Pereira (more urban)

Agrupamento de Escolas Gabriel Pereira

Rough distribution of students


```
1  student_both.G3_x[student_both.address =="U"].mean()
```
10.724137931034482

```
1  student_both.G3_x[student_both.address =="R"].mean()
```
9.555555555555555

PROPOSAL:
- Perhaps one of the students painpoints is having an urbanised environment to learn the best that they can
- Encourage government to build more towns

Sub-question based on ideals

- How can the government/school create policies
  - to help students achieve their ideal goals?

- Why?
  - So that government/schools can ==improve the overall quality of education== with a more multi-pronged, holistic approach instead of just one single approach

Ideals

- Things that students would value as it would bring them ==convenience/entertainment==

- **traveltime**
  - home to school travel time
  - (numeric: 1 – <15 min., 2 – 15 to 30 min., 3 – 30 min. to 1 hour, or 4 – >1 hour)

- **studytime**
  - weekly study time
    - (numeric: 1 – <2 hours, 2 – 2 to 5 hours, 3 – 5 to 10 hours, or 4 – >10 hours)

- **freetime**
  - free time after school (numeric: from 1 - very low to 5 - very high)

- **goout**
  - going out with friends (numeric: from 1 - very low to 5 - very high)

Ideals

- internet
  - Internet access at home (binary: yes or no)

- romantic
  - with a romantic relationship (binary: yes or no)

- higher
  - wants to take higher education (binary: yes or no)

- activities
  - extra-curricular activities (binary: yes or no)

- reason
  - reason to choose this school (nominal: close to 'home', school 'reputation', 'course' preference or 'other')

Ideals (factors)

- ## G1/G2/G3
  - first period grade (numeric: from 0 - 20)

- ## failures
  - number of past class failures (numeric: n if 1<=n<3, else 4)

- ## absences
  - number of school absences (numeric: from 0 to 93)

suffix
_x is for math
_y is for portugese)

Ideals (factors): Null Hypothesis

```python
# T-test to check if traveltime affects traveltime
Ho = "Traveltime affects grades negatively"        # Stating the Null Hypothesis
Ha = "Traveltime doesnt affect grades negatively"   # Stating the Alternate Hypothesis

x2 = np.array(student_both.G3_x[student_both.traveltime == 4][student_both.G3_x<10])
# Selecting grades for travel time == 4 as we saw in the graph it is a downward trend
y2 = np.array(student_both.G3_x)
# Selecting comparison to ALL grades for G3 (which is dependent on G1 & G2)

t, p_value  = stats.ttest_ind(x2,y2, axis = 0)   #Performing an Independent t-test

if p_value < 0.05:   # Setting our significance level at 5%
```

Traveltime affects grades negatively as the p_value (0.05507671760061295) > 0.05,
 Alt hypothesis is not statistically significant and indicates,
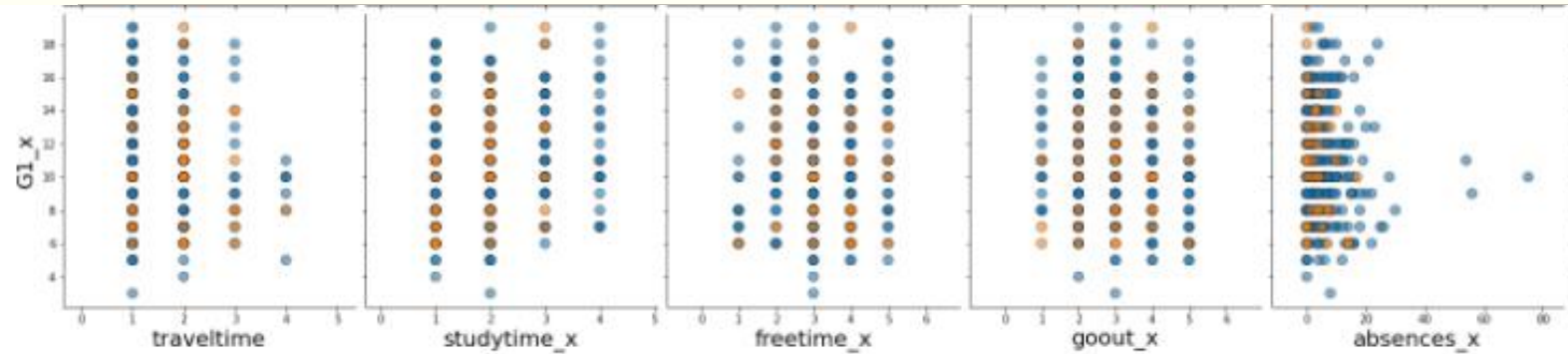 strong evidence for the null hypothesis as there is more than 95% probability the null is correct.

Ideals (factors): Grades Correlation

Eg. For every increment in grades (0-20) we get a roughly 25% increment in study time for Portugese. In general very weak +/-ve correlation

	G3_x	G2_x	G1_x	traveltime	studytime_x	freetime_x	goout_x	failures_x	absences_x
G3_x	1.000000	0.906681	0.805134	-0.131892	0.074705	0.006605	-0.124181	-0.373315	0.018794

	G3_y	G2_y	G1_y	traveltime	studytime_y	freetime_y	goout_y	failures_y	absences_y
G3_y	1.000000	0.889678	0.816962	-0.153028	0.272652	-0.106566	-0.135305	-0.340582	-0.119905

Travel time > 1 hour
Grades

More studytime

Min. Grades improve slightly at 5 – 10 or more hours
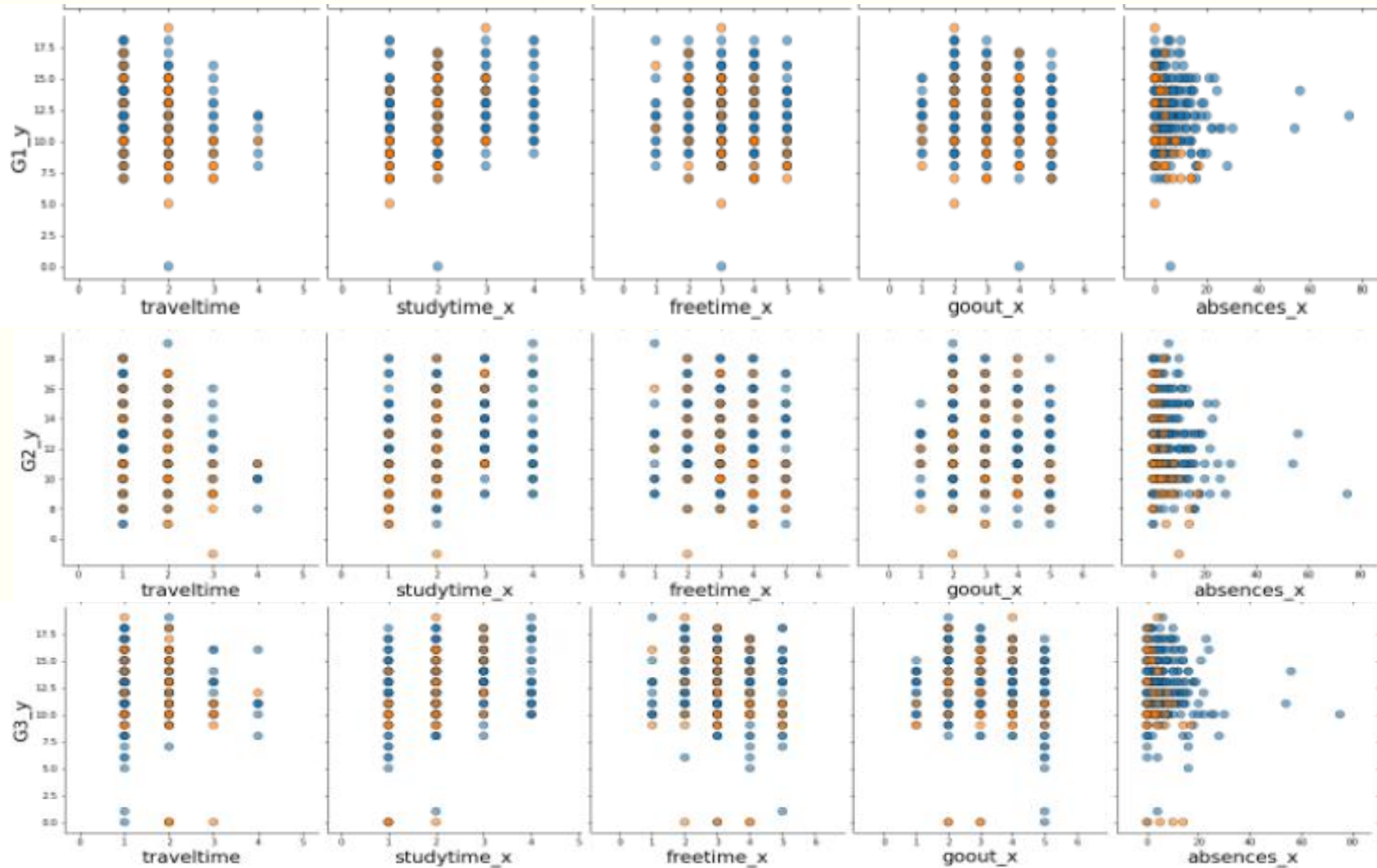
More freetime

No effect

Going out more

No effect

absenteeism
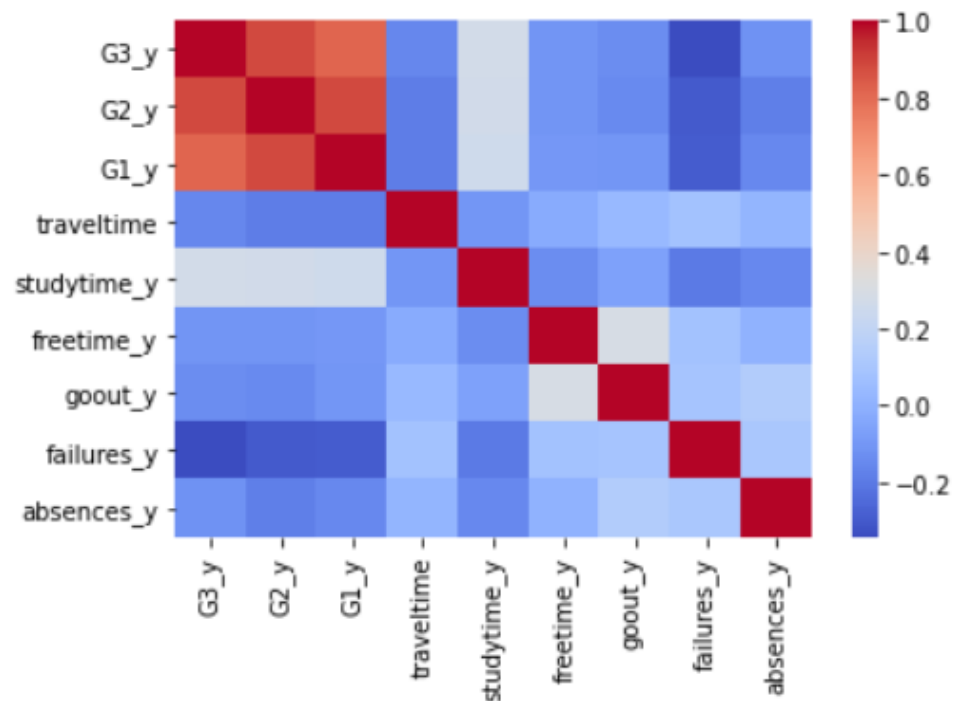
No effect
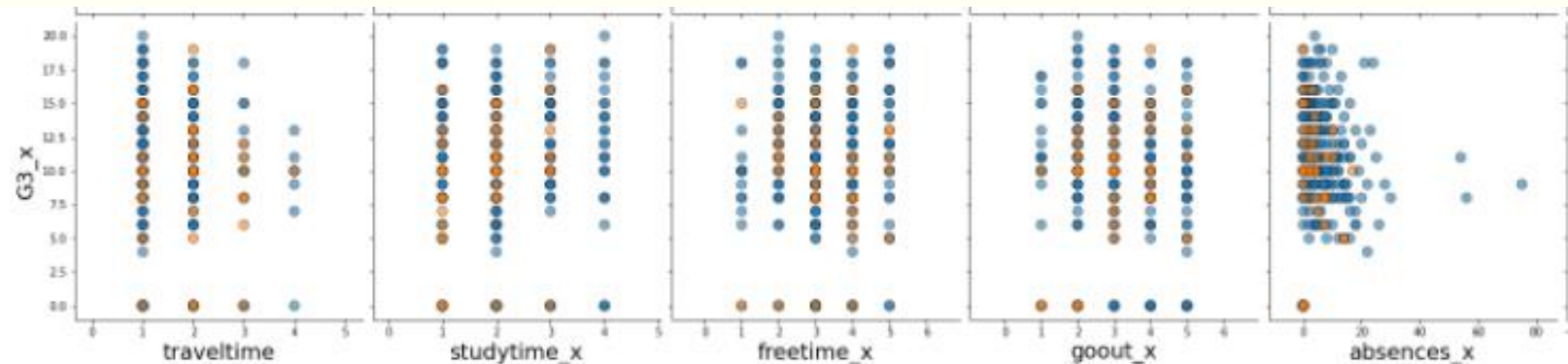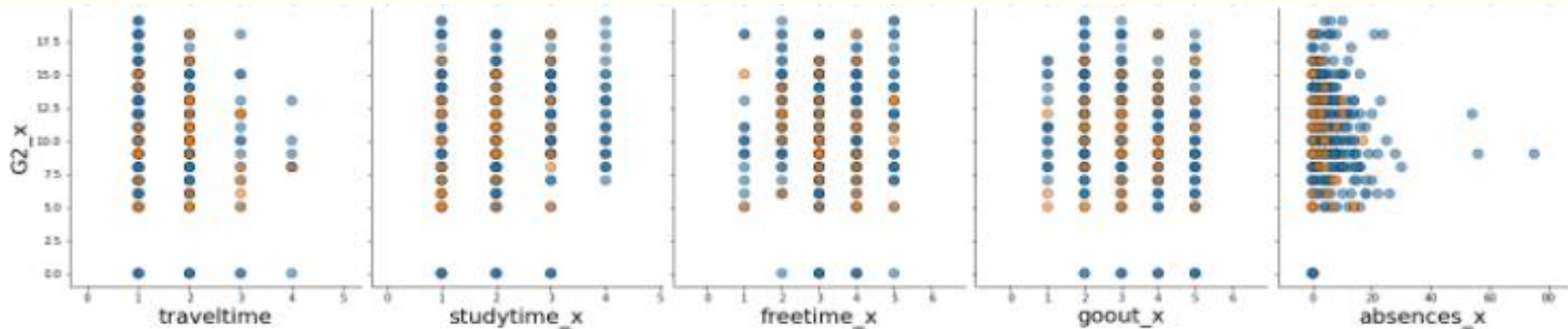
How ideals can affect Grades: PORTUGESE (PairPlots)

Travel time > 1 hour
Grades ⬇

More studytime

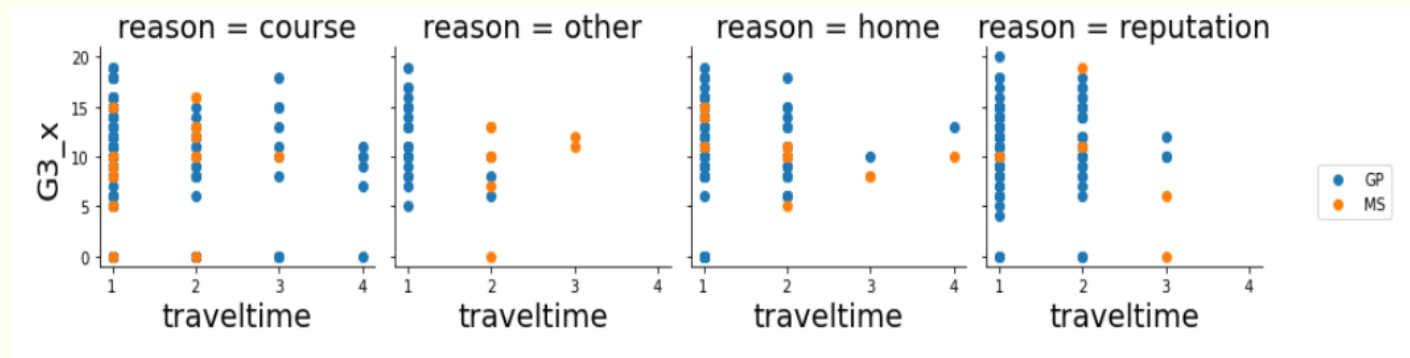Min. Grades improve slightly overall

More freetime

No effect /inconclusive

Going out more

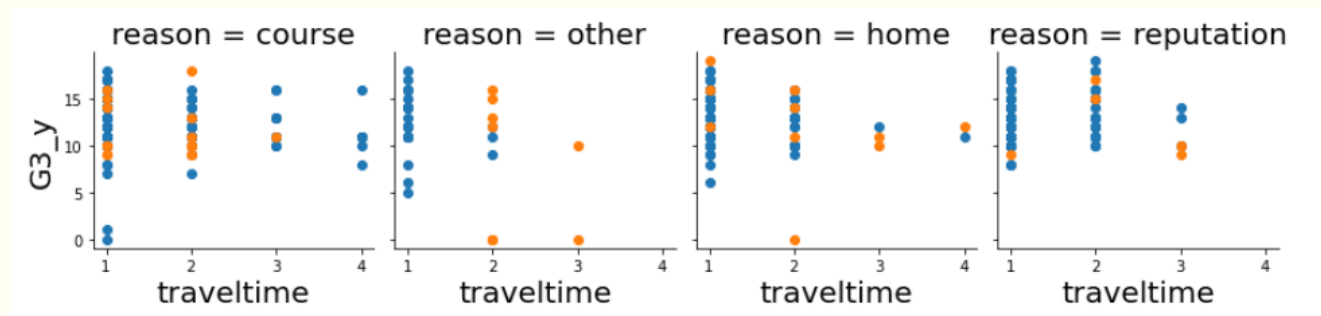(At 5) too much time with friends decreases grades

More absenteeism

No effect

# How ideals(Traveltime & Reason) can affect Grades (FacetGrid)

- We can see travel time exceeds an hour when the desired course/traveltime is only located in a certain school.

- when travel exceeds an hour, grades drop

Insights from Grades (traveltime)

- - Travel time exceeding > 1 hour statistically affects grades:
  - Null hypothesis test, weak negative correlation & visually seen via scatter plots
    - Grades dropped sharply at the > 1 hour travelling time mark
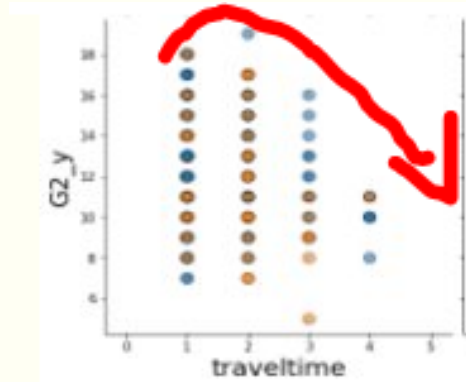
- Recommendation to client:
  - Based on reason: perhaps
  - based on course (travel time increases) and home (travel time decreases)
    - Can provide financial assistance similar to grants for rent near school (eg. Halls in local uni in Singapore)


- Now that is is COVID-19 period traveltime will naturally drop.
- PERHAPS a new dataset is required to further explore how a home based setting can affect grades, absenteeism and failures,Now that is COVID-19 period traveltime will naturally drop.
  - PERHAPS a new dataset is required to further explore how a home based setting can affect grades, absenteeism and failures,
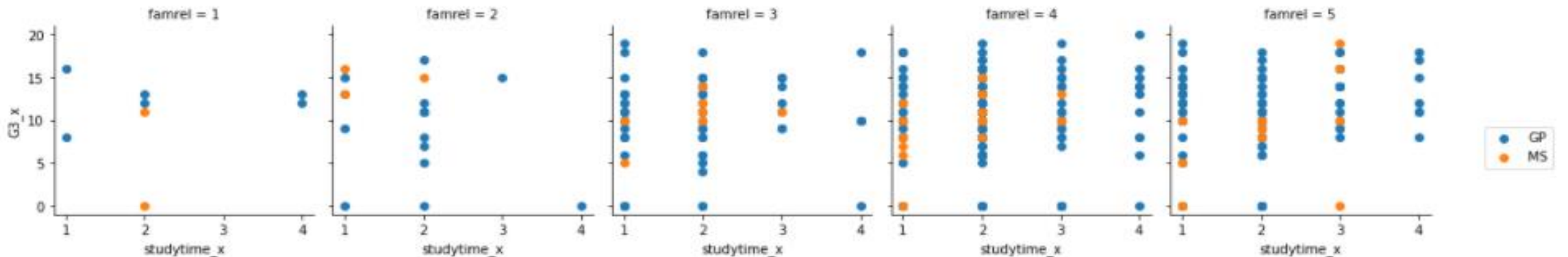
- Proposal:
  - If confident that traveltime drop will improve overall grades can continue with School from home even after COVID-19 pandemic is over

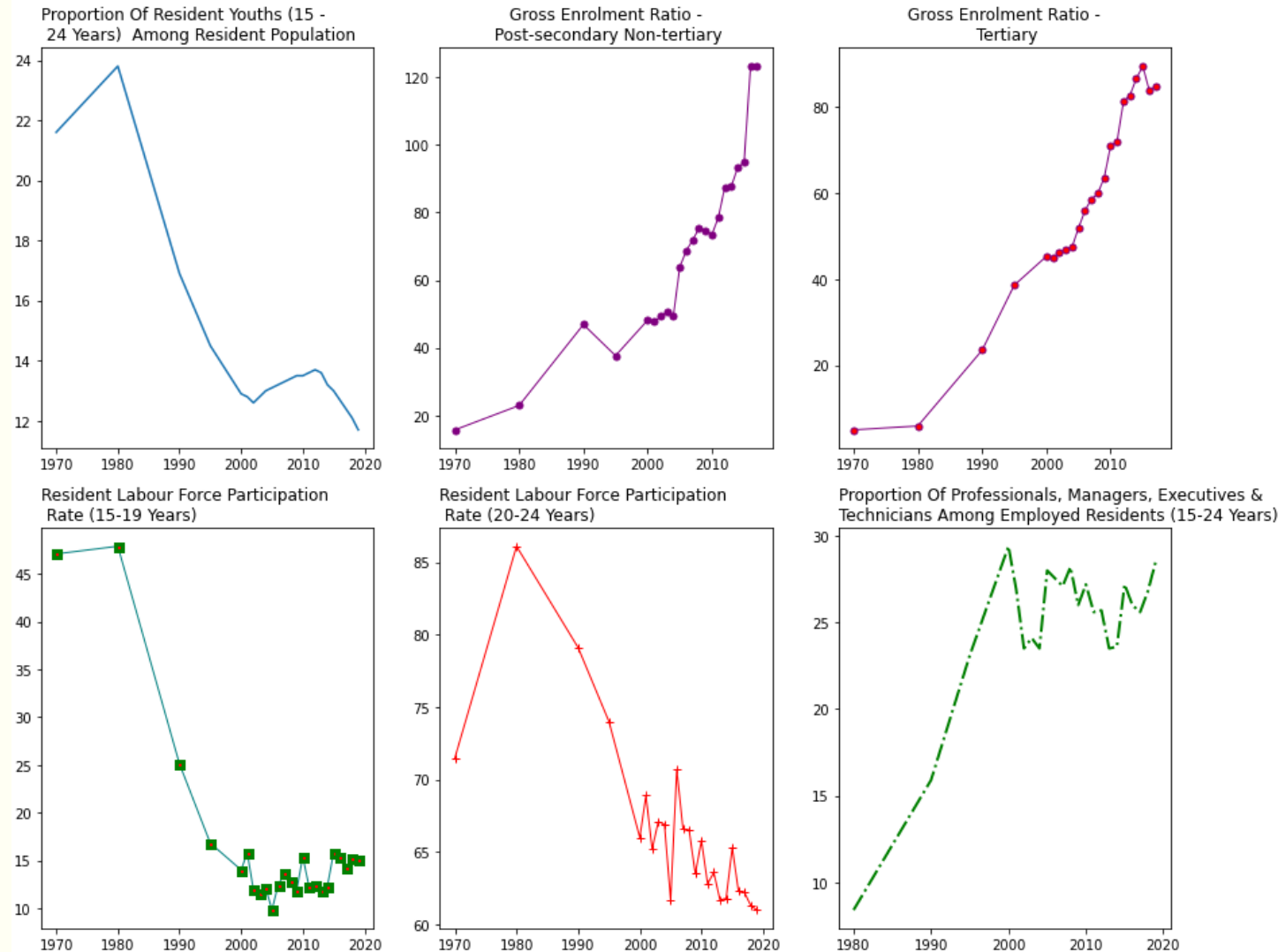   or a mix of off & online education

How ideals(studytime) can affect Grades (an overview)
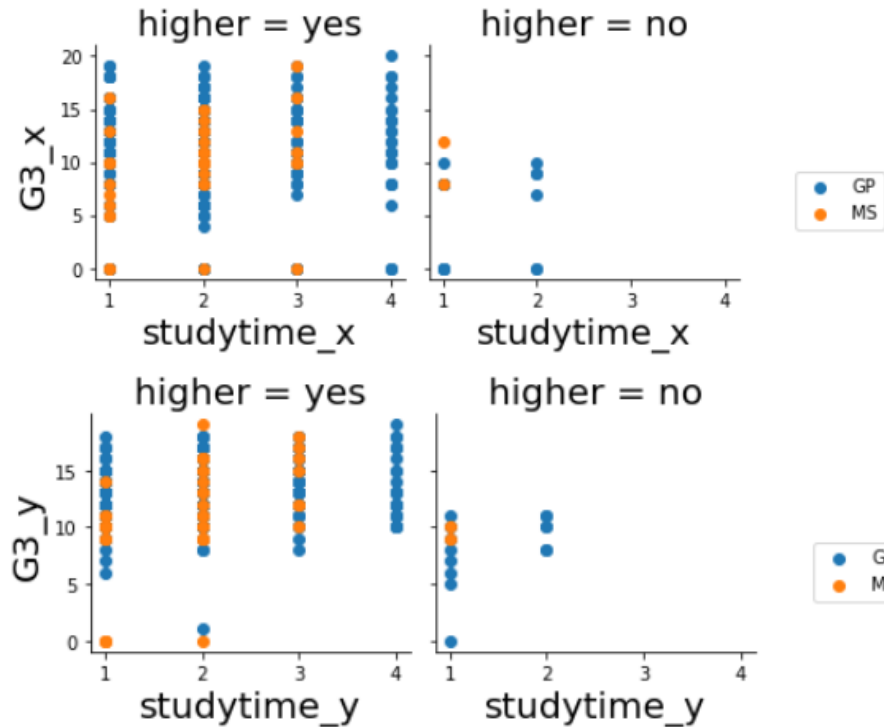
- Most students are motivated to take on higher education
- Those who are not have lower study time and a lower mean grade.
- A lower study timemight cause them to score lower marks than normal
- Typically more studytime impacts mental health, lets take FAMREL as a proxy for that.
- The healthier the famrel, the more study time students get plateauing at 4 & 5

Insights from Grades (Studytime) + SG context/dataset

Proportion Of Resident Youths (15 - 24 Years) Among Resident Population

Gross Enrolment Ratio - Post-secondary Non-tertiary

Gross Enrolment Ratio - Tertiary

Resident Labour Force Participation Rate (15-19 Years)

Resident Labour Force Participation Rate (20-24 Years)

Proportion Of Professionals, Managers, Executives & Technicians Among Employed Residents (15-24 Years)

1. Take post secondary enrolment as a proxy for "Higher" in our main dataset

2. Take lower working teens + more professional jobs (15-24) as a sign of a healthier lifestyle

PROPOSAL:
- Since Singapore is already doing very well in promoting higher education (which translates to better jobs)
- We can do better by improving mental health
- When work gets too stressed singaporeans get more and more reluctant to marry, let alone have a family which is shown in the very low % of youths currently
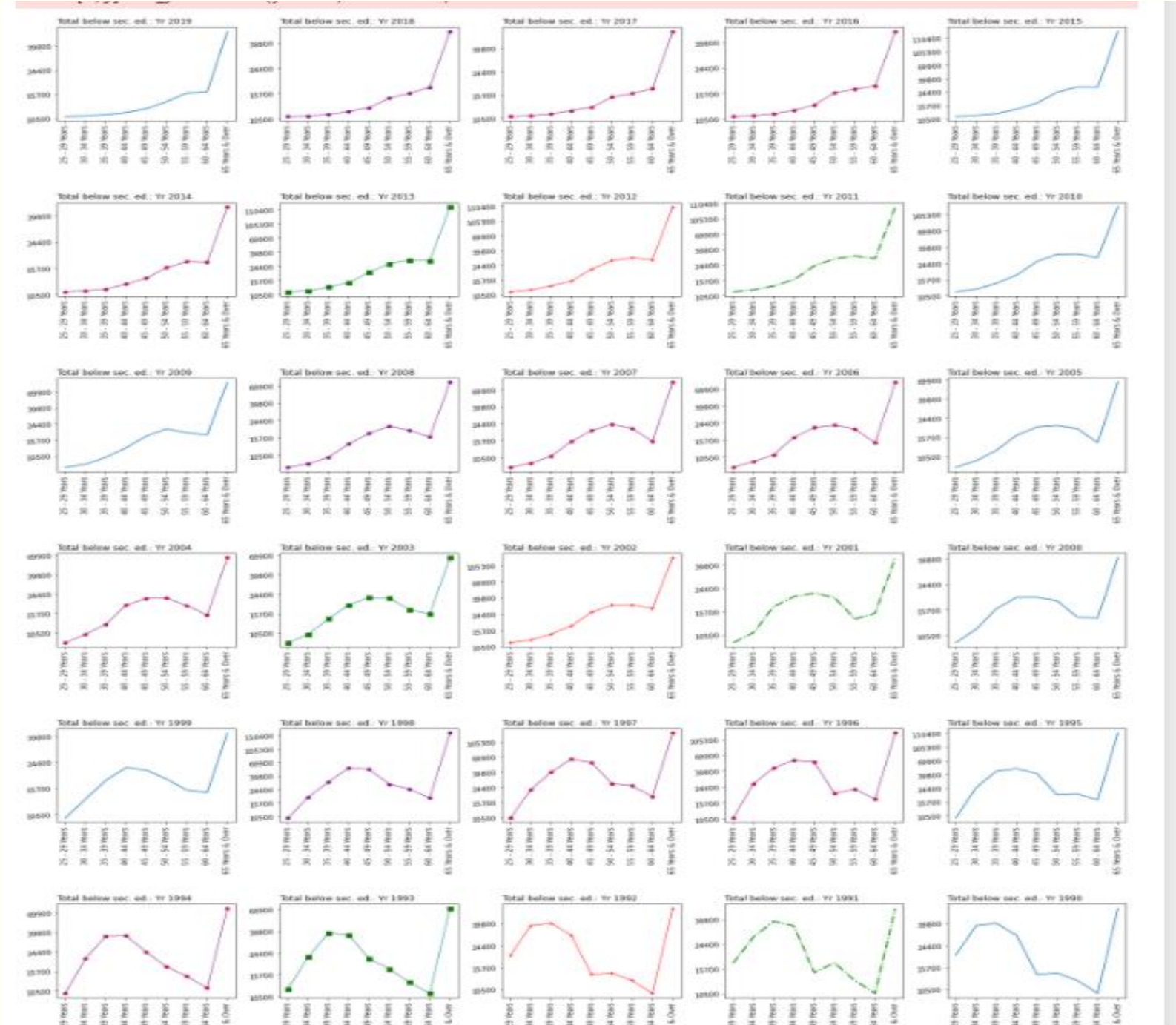
viewing over 30 years how the youths decreased to a low%

News feed

28 Jul 2020 10:45AM
(Updated: 28 Jul 2020 10:39PM)

Bookmark

Singapore

Fewer marriages, more divorces in Singapore last year

Fewer couples in Singapore got married in 2019. There was also an uptick in divorces. A Department of Statistics report reveals the number of babies born last year also increased slightly, after an eight-year low in 2018. CNA's Vanessa Lim reports.

SINGAPORE: Fewer couples in Singapore got married and more marriages ended in divorce last year, the Singapore Department of Statistics (SingStat) said on Tuesday (Jul 28).

THE STRAITS TIMES

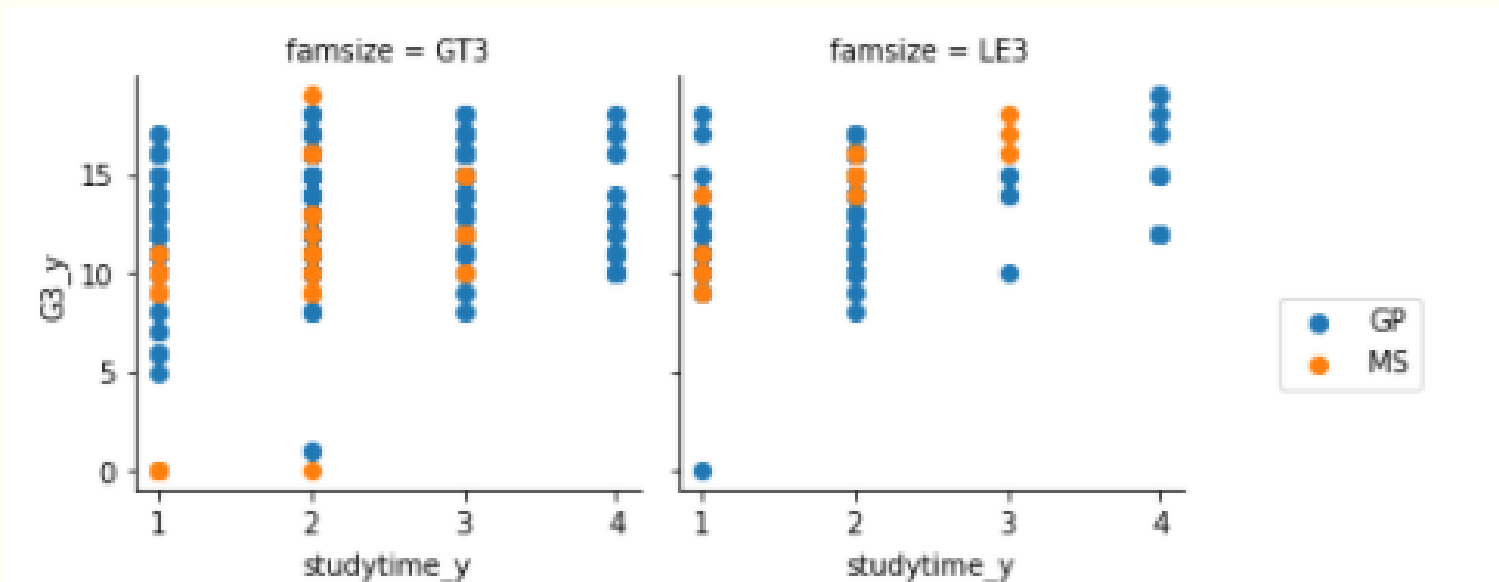SINGAPORE   LOG IN

More women in Singapore are staying childless

Changing values, career focus reasons for rate nearly tripling; experts see trend continuing

The proportion of women in Singapore who stay childless has almost tripled in the past 20 years. ST FILE PHOTO

Grades & Famsize

While we want to encourage Singaporeans to have children.

Our dataset actually suggests that Family sizes >3 actually have a lower average grade score for the final score

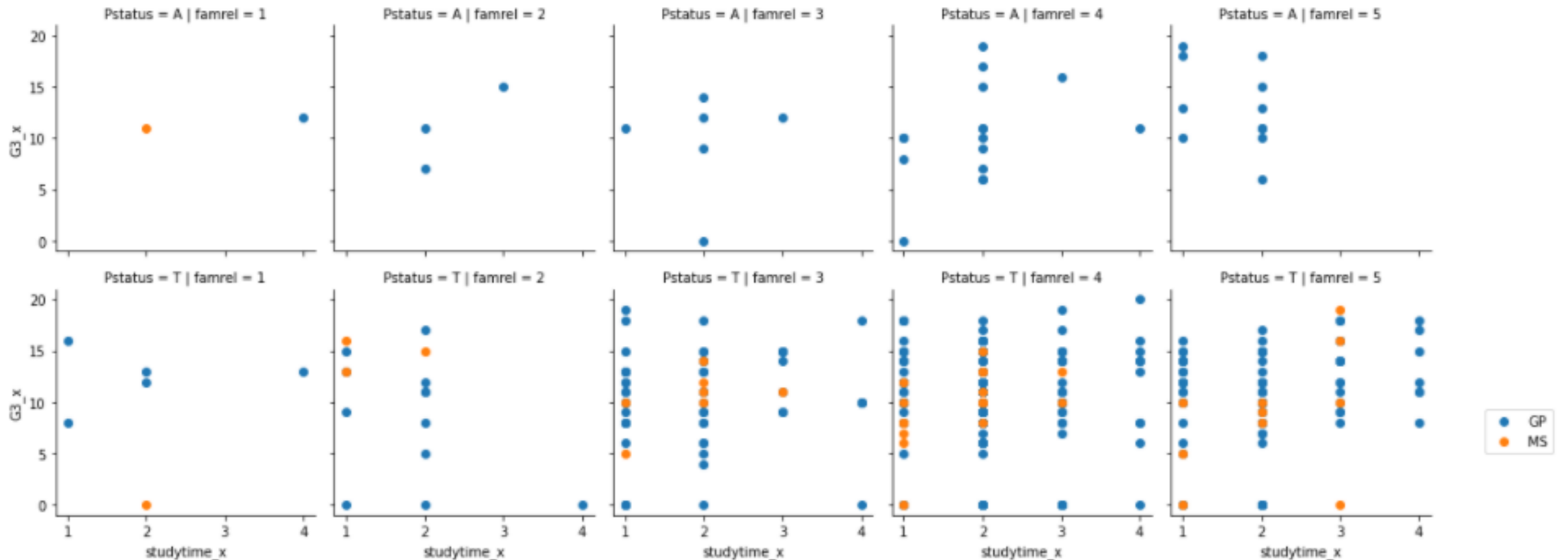So while we would want to encourage

```
1  student_both.G3_x[student_both["famsize"]=="LE3"].mean()
```

11.23076923076923

```
1  student_both.G3_x[student_both["famsize"]=="GT3"].mean()
```

10.172284644194757

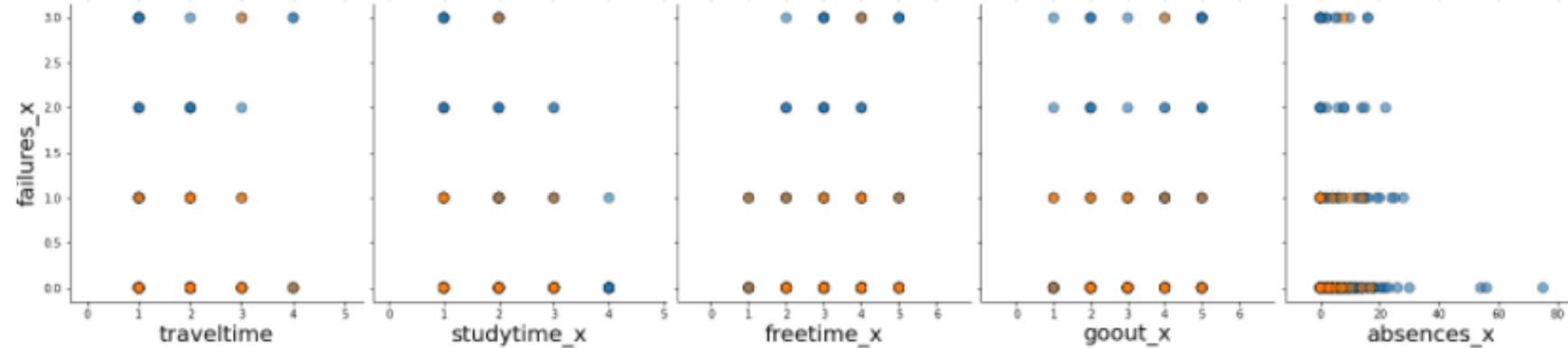# Grades, famrel, Pstatus & studytime (a glance back into dataset)

True enough, while pursuing pure academic results, we indirectly lose KPIs in academic factors (eg. Studytime)

No matter how good the famrel…. if parents are apart study time only reaches >10 hours once

Travel time

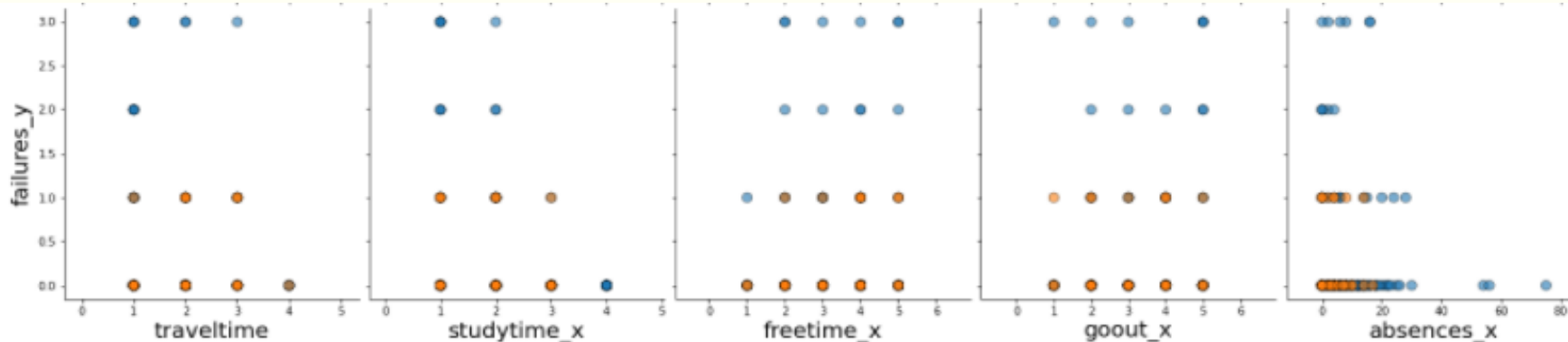No effect

More studytime

Failures

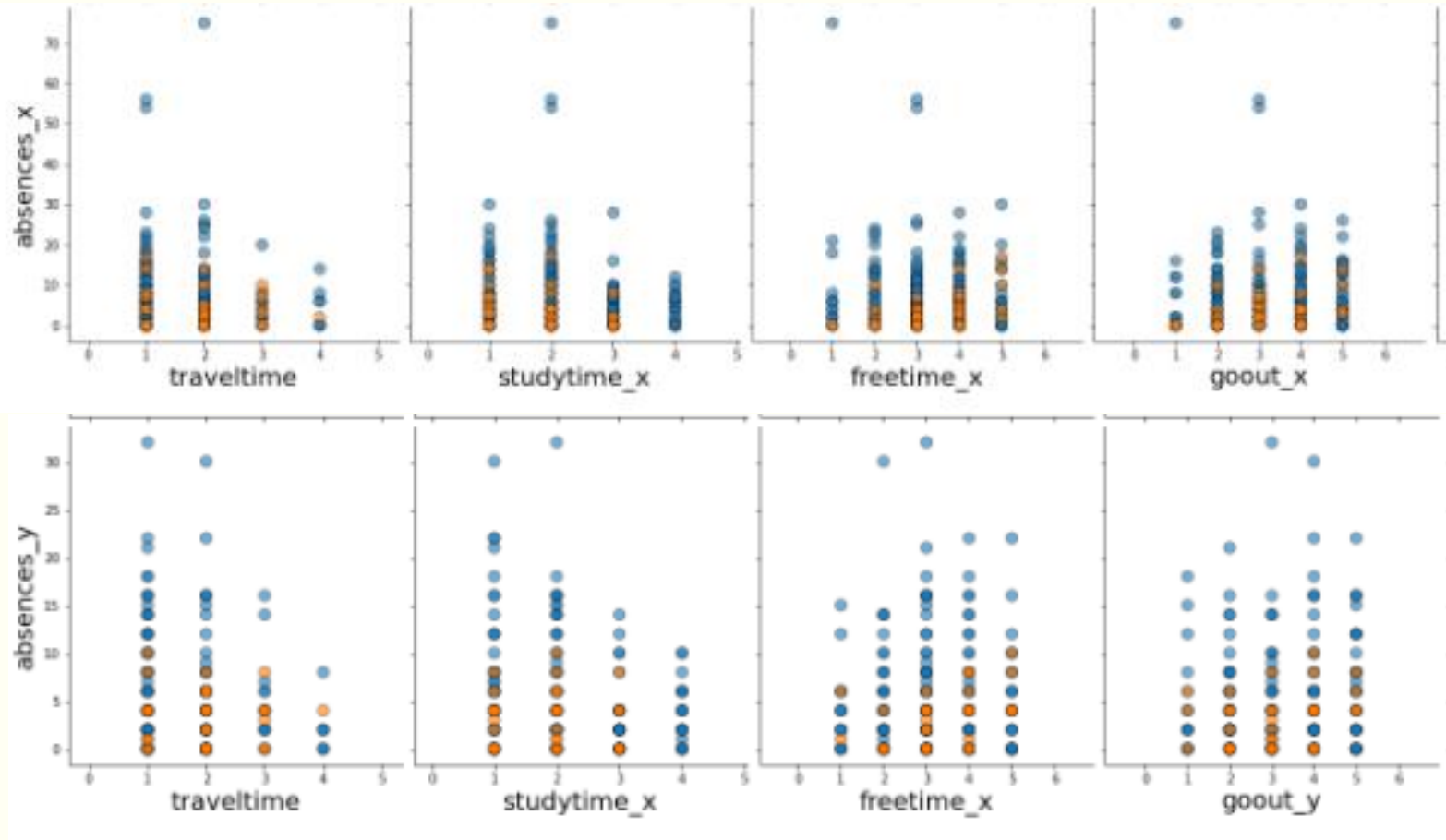More freetime

No effect

Going out more

No effect

More absenteeism??

It seems like more absenteeism drops failures
Which doesn't make sense
Spurious finding perhaps

How ideals can affect Absences: MATH &PORT (an overview)

Insights on long travel time = low absenteeism

MATH

PORTUGESE

Grades
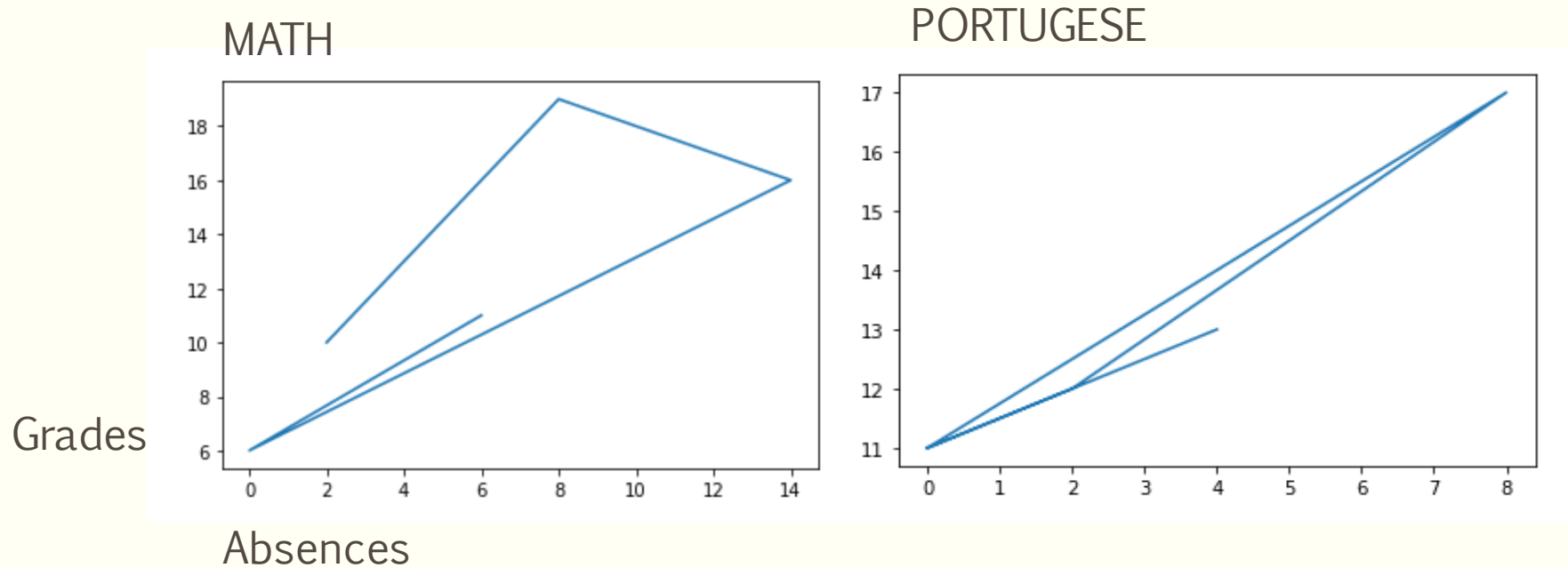
Absences

Conclusion:
while it is true that a longer travel
time = lower absenteeism, it
doesn't mean we should encourage
a long travel time due to earlier
findings and we should take it as a
coincidence

|	G3_x	G2_x	G1_x	traveltime	studytime_x	freetime_x	goout_x	failures_x
absences_x	0.018794	-0.048014	-0.060171	-0.017870	-0.073389	-0.045319	0.073656	0.033922

Internal/External support (factors)

- Paid tuition /External govt. Or school assistance
- **famsup**
  - family educational support (binary: yes or no)

- **schoolsup**
  - extra educational support (binary: yes or no)
- **paid**
  - extra paid classes within the course subject (Math or Portuguese) (binary: yes or no)

Internal/external support doesnt affect grades: H0 (null hypo)

```python
# T-test to check if traveltime affects traveltime
Ho = "ext./int assistance doesnt affect grades"        # Stating the Null Hypothesis
Ha = "ext./int assistance affects grades"              # Stating the Alternate Hypothesis

x2 = np.array(student_both.G3_x[student_both.paid_x == "no"][student_both.famsup_x == "no"]\
              [student_both.schoolsup_x == "no"])
# students who have NO support
y2 = np.array(student_both.G3_x[student_both.paid_x == "yes"][student_both.famsup_x == "yes"]\
              [student_both.schoolsup_x == "yes"][student_both.G3_x>10])
# students who have ALL support and at least passed

t, p_value  = stats.ttest_ind(x2,y2, axis = 0)   #Performing an Independent t-test

if p_value < 0.05:  # Setting our significance level at 5%
```

ext./int assistance doesnt affect grades as the p_value (0.4794973401464786) > 0.05,
 Alt hypothesis is not statistically significant and indicates,
 strong evidence for the null hypothesis as there is more than 95% probability the null is correct.

Internal/External support
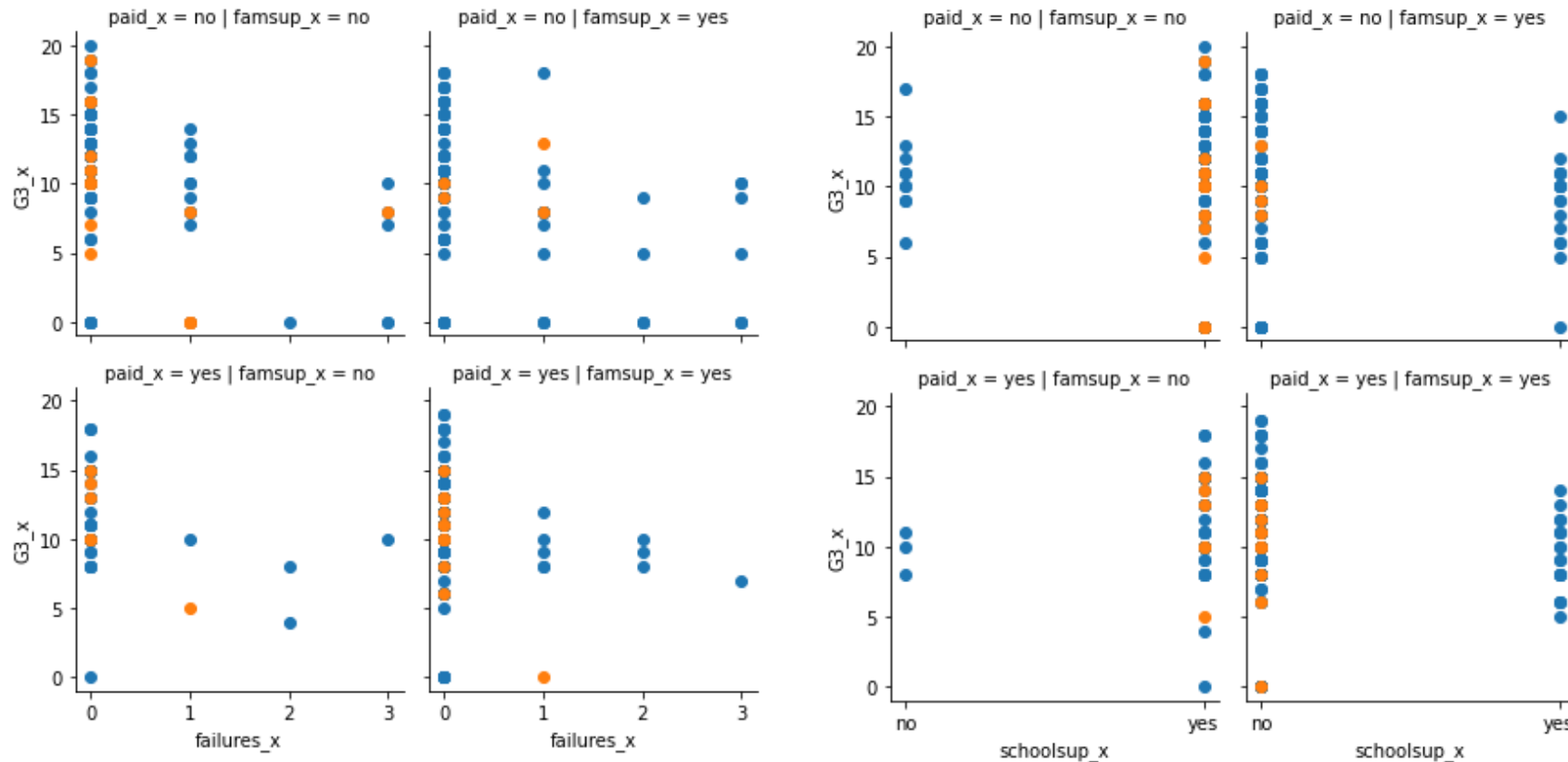
No conclusive insights…

We can say that support by family/school/other has no significant impact on grades

While it is a statistical finding perhaps there are more reasons why even with all the support going into education…it doesn't give us a clear trend

PROPOSAL: we shouldn't recommend less investment in education, but perhaps find out more underlying reasons why the input in education is not working as what is logically supposed to occur.

End. Fin.

- https://github.com/lolasery/khdatasci

- Ty!