School Absence?

Cyzus Chi

Overall Motivation

Variables

Absence: number of school absences

Traveltime: time that student needs to take to get to school

Age: the age of students

Health: health status of students

Internet: whether or not having Internet access at home

Medu & Fedu: Mother education & father education

#1 Does age affect the number of absences?

Students are becoming…

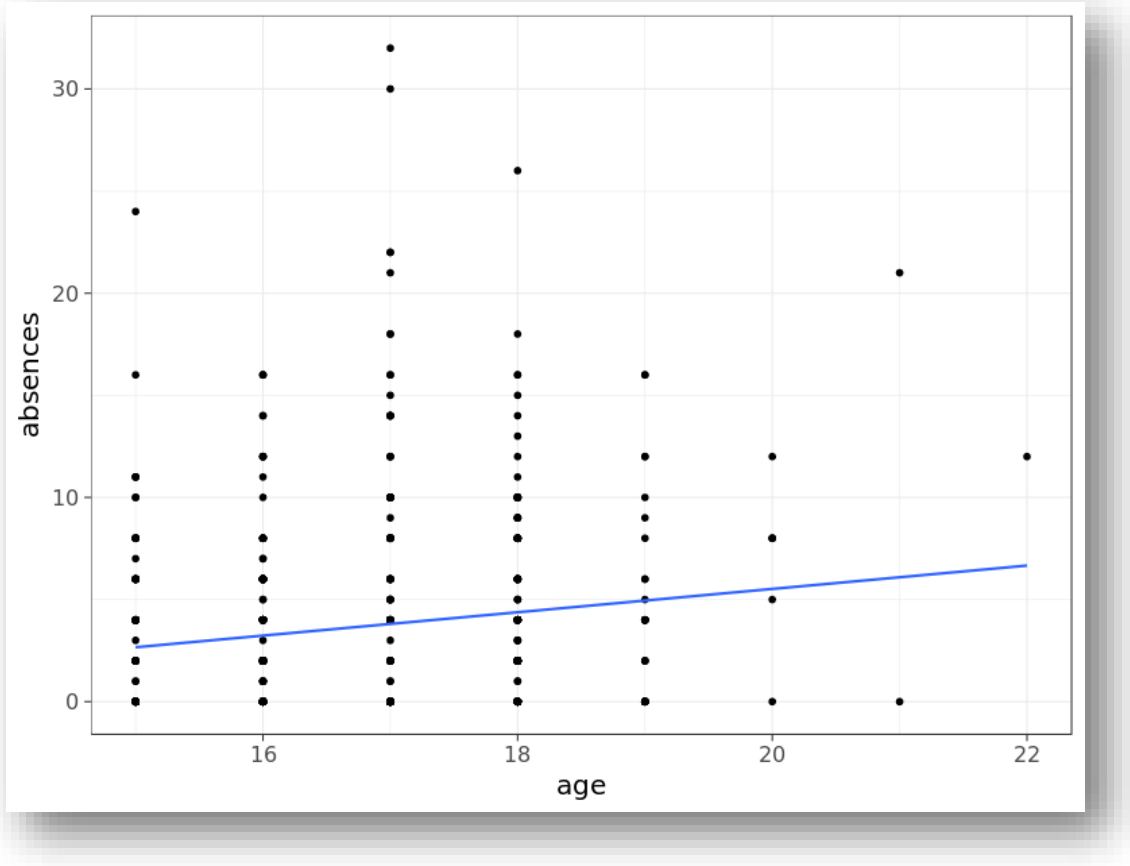
1. Busier
2. More rebellious…?

<u>Motivation</u>

As students become older, it is reasonable that they become busier or more rebellious. School attendance may not be that important for they have other work to do or they simply do not want to go to school

#1 Does age affect the number of absences?

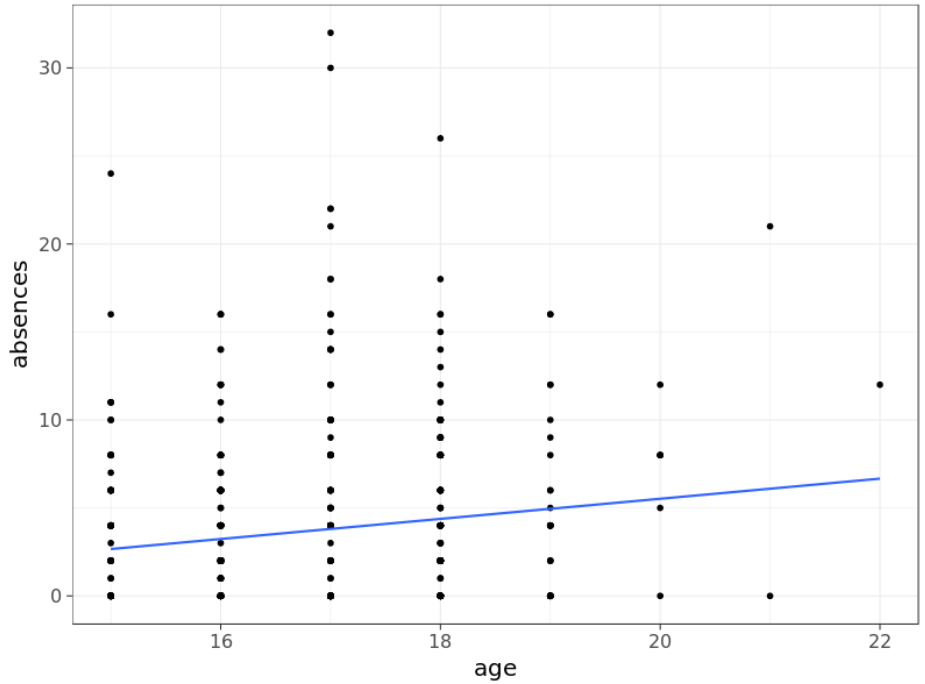
- **Question**: Is there a correlation between age and absence?
- **Assumption**: As students become older, they are tend to have a higher number of absences
- **Null:** The correlation coefficient between age and absences is 0
- **Alternate:** The correlation coefficient is greater than 0



#1 Does age affect the number of absences?

$$H_0: \rho = 0 \; VS \; H_A: \rho > 0$$

```
pValue: 0.0001
statistic: 3.859
95% confidence interval: [0.0862,1]
```



<u>Conclusion</u>

Because p-Value is smaller than the significance level α(0.05), we can reasonably reject null hypothesis. According to 95% CI, the real correlation could be within [0.0862,1]. The sample correlation is 0.150, it might be due to the potential outliers (the age of 20+ students). Therefore, we have sufficient evidence to claim that there is a correlation between age and absence greater than 0.

#2 Does travel time associate with absence?

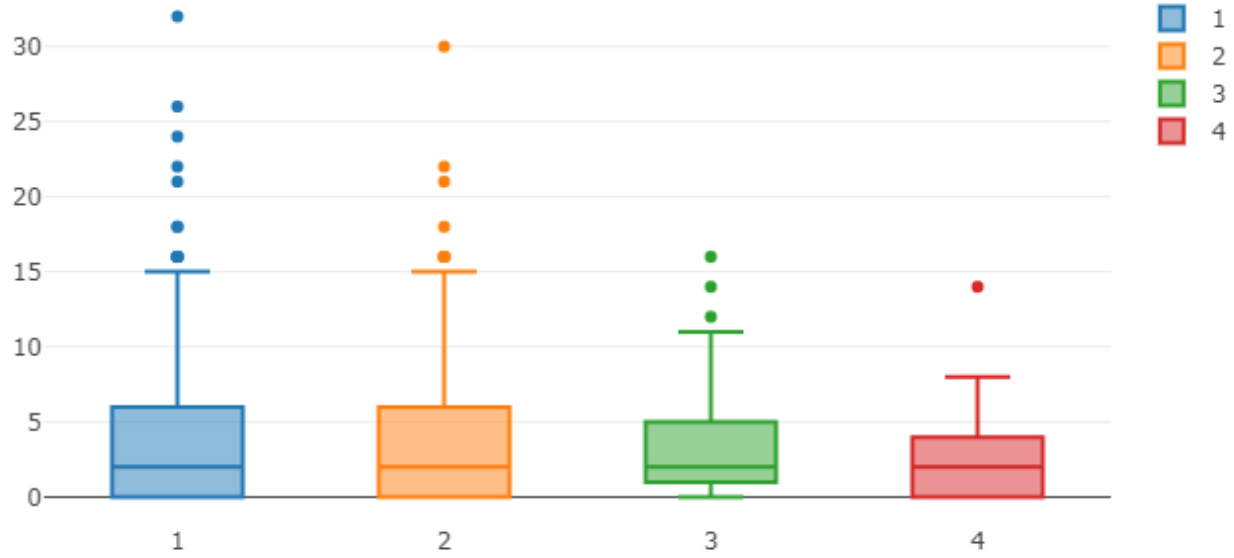
<u>Motivation</u>

The distance between school and home differs from student to student. There are many possible events such as traffic jams or accident injury happened when students are going to school. So is that possible that students who have a longer travel time to school might have a greater number of absences?

#2 Does travel time associate with absence?

- **Question**: Does different travel time of students have a different mean number of absences?

1: < 15
2: 15-30
3: 30-60
4: >60



#2 Does travel time associate with absence?

***Question**: Does different travel time of students have a different mean number of absences?*

Hypothesis:

- **Null:** The mean number of absence for students with different travel time are all the same
- **Alternate:** The mean number of absence for students with different travel time are not all the same

1 – <15 min
2 – 15-30 min
3 – 30-60 min
4 – >60 min

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$$

$$H_A: not\ (\mu_1 = \mu_2 = \mu_3 = \mu_4)$$

#2 Does travel time associate with absence?

```
Analysis of Variance Table

Response: absences
             Df  Sum Sq Mean Sq F value Pr(>F)
traveltime    3     7.7  2.5713  0.1189  0.949
Residuals   645 13948.0 21.6249
```

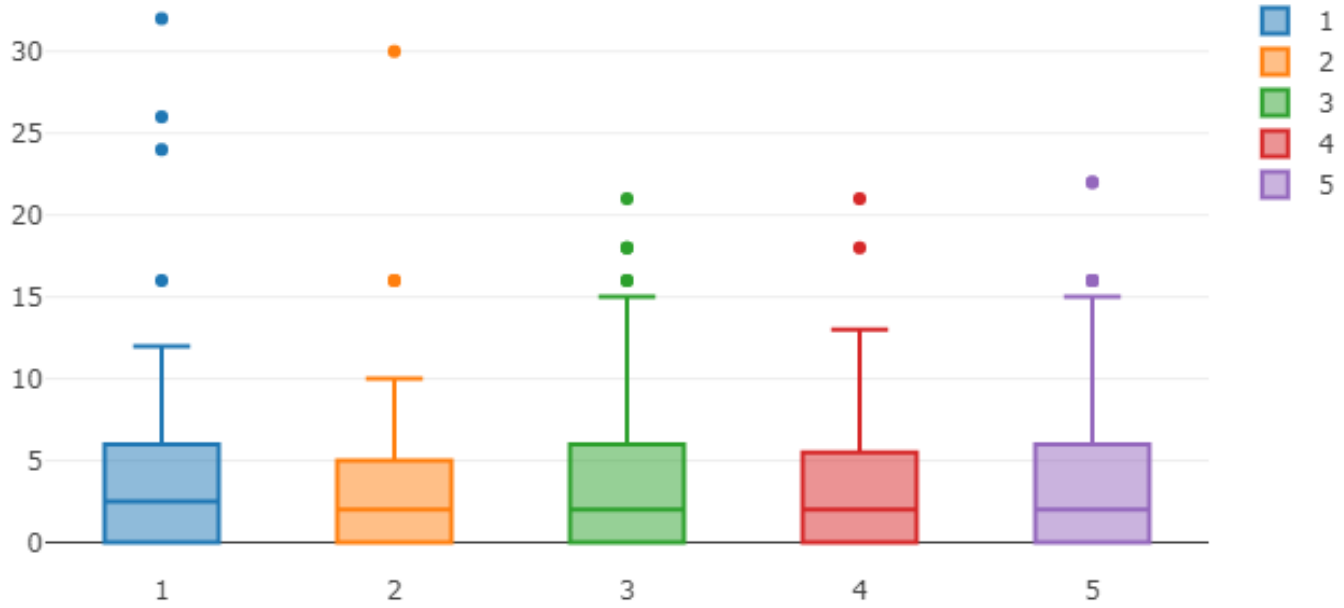
The result is actually surprising. F-value is relatively larger than what we think of. Since p-value is greater than the significance level of 0.05, we do not reject null hypothesis. In context, higher or lower traveltime doesn't necessarily mean that those students would have a higher or lower mean number of absence

#3 What about health status?

Motivation

It is usual to catch a cold or get sick during the semester. Mostly when we are terribly sick, we might simply not attend the class and stay in home to get recovered.

#3 What about health status?



1 – very bad
5 – very good

#3 What about health status?

***Question**: Does poor health status associate with a higher school absence*

Hypothesis:

- **Null:** The mean number of absence for students with different health status are all the same
- **Alternate:** The mean number of absence for students with different health status are not all the same

1 – very bad
5 – very good

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$$

$$H_A: not\ (\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5)$$

#3 What about health status?

```
Analysis of Variance Table

Response: absences
           Df  Sum Sq Mean Sq F value Pr(>F)
gender      1     6.4  6.3529  0.2947 0.5874
Residuals 647 13949.4 21.5601
```

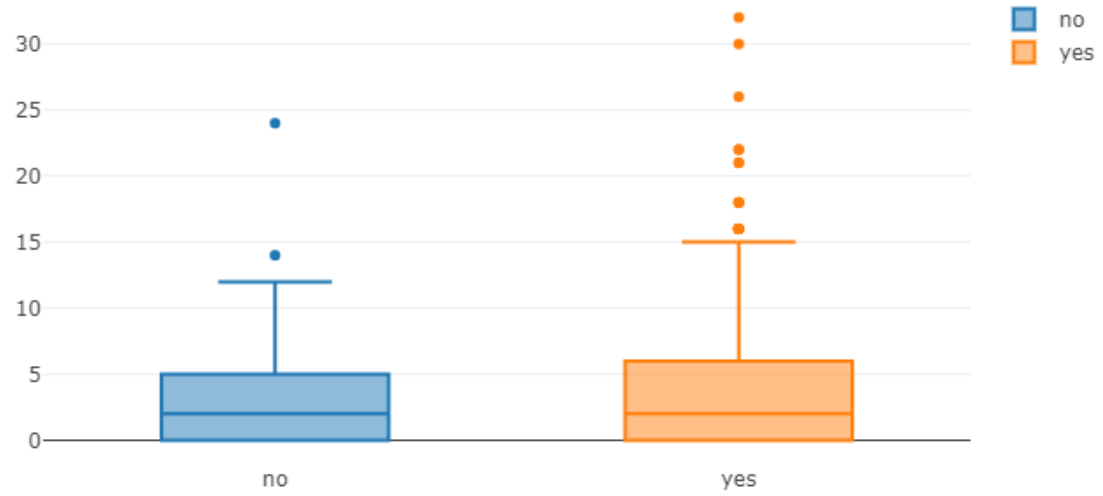
F-value is 0.2947. Since p-value is greater than the significance level of 0.05, we do not reject null hypothesis. This means that we don't have sufficient evidence to tell that the mean absence for different health status are not the same.

#4 Does Internet access predict absence?

<u>Motivation</u>

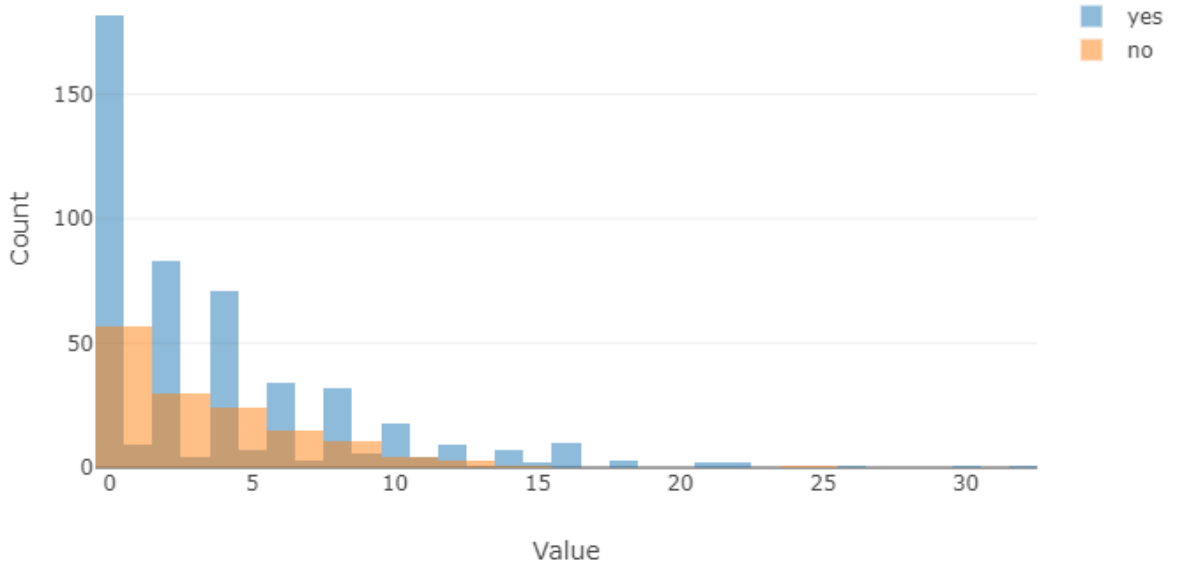
Internet may be a considerable distraction to students. Could we be find any evidence showing that students with Internet access at home have a greater mean of absence?

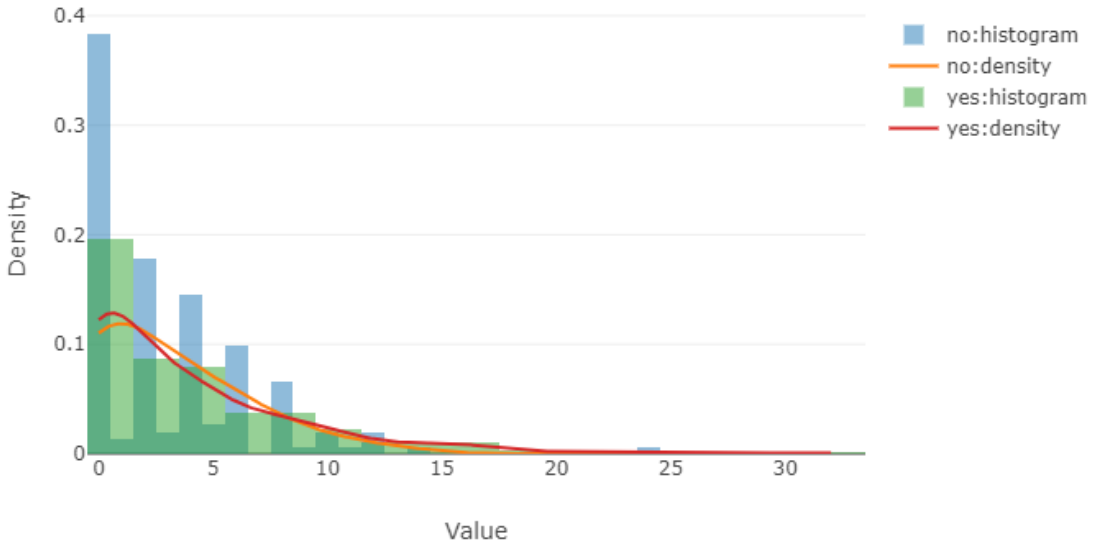
#4 Does Internet access predict absence?



According to the box plot, IQR of two groups mostly overlap

#4 Does Internet predict absence?





Students with no Internet access at home seem to have lower number of absence
Considering the total number differs, I also draw a density curve

#4 Does Internet access predict absence?

$$H_0 : \mu_{\text{internet:no}} - \mu_{\text{internet:yes}} = 0$$

$$H_1 : \mu_{\text{internet:no}} - \mu_{\text{internet:yes}} \neq 0$$

```
pValue: 0.0464
statistic: -1.9913
95% confidence interval: [-1.4656,-0.0116]
```

The p-value is less than the significance level of 0.05, so we reject null hypothesis. It is 95% confident to tell the actual difference can be within the interval [-1.4656,-0.116]. Therefore, in this case, the data might suggest that students with Internet access might have a greater absence mean.

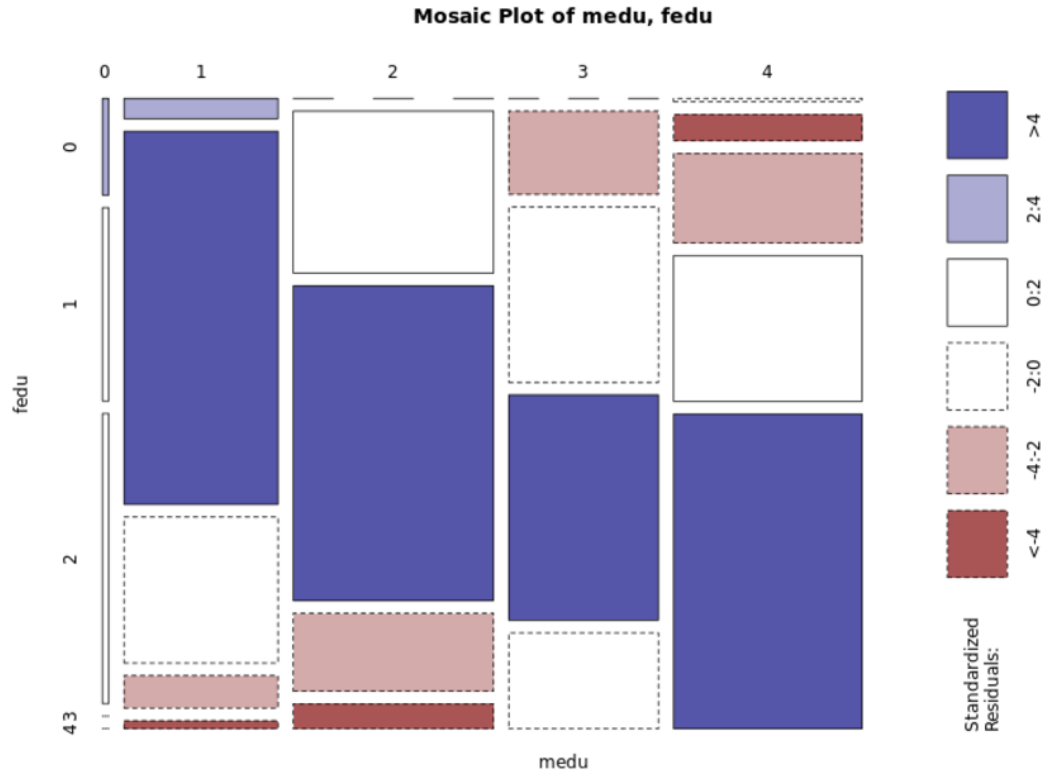
#5 Does medu & fedu affect absence?

<u>Motivation</u>

We have good reasons to believe that parents play an important role in students' education. So looking at the parents' educations might be a good idea.

#5 Does medu & fedu affect absence?

Mosaic Plot of medu,fedu:



It can be told from the mosaic plot that the students' mdu and fdu are not independent. Instead, a student's father with a high education tends to have a mother with high education as well and vice versa.

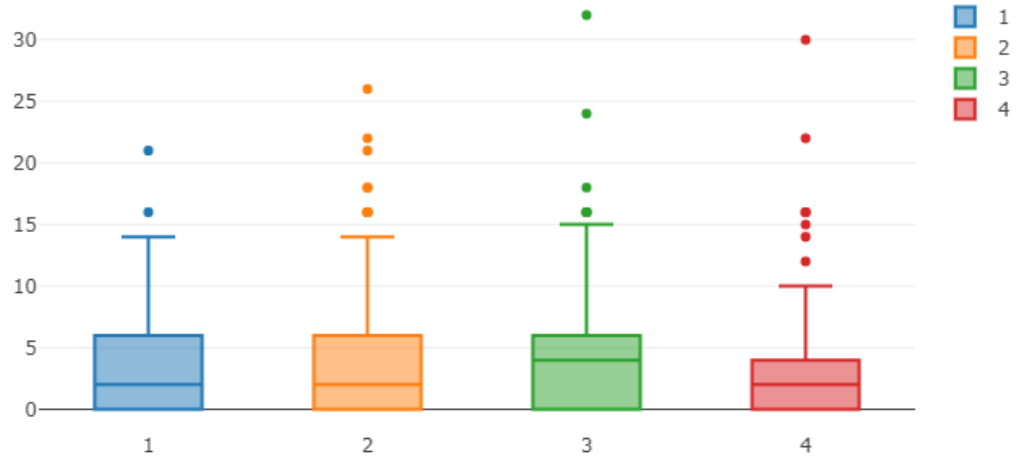
#5 Do medu & fedu affect absence?

Initially, the data set includes the moms and dads with no education, but the quantity of these categories is too small.

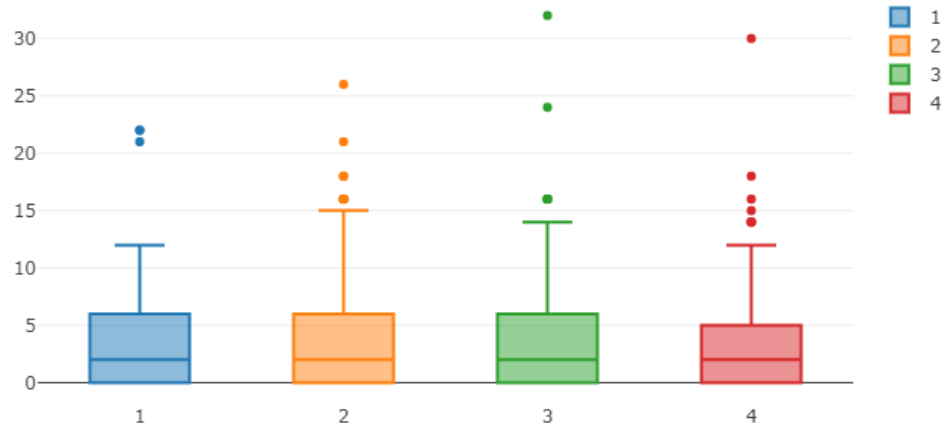
Absence grouped by medu

Absence grouped by fedu

#5 Do medu & fedu affect absence?



Absence grouped by medu



Absence grouped by fedu

Generally, their 75 percentiles align with number of 5

#5 Do medu & fedu affect absence?

***Question**:* Do medu & fedu *associate with their children's school absence?*

Hypothesis:

- **Null:** The mean absence for students whose parents have different education levels are the same
- **Alternate:** The mean absence for students whose parents have different education levels are not all the same

1 – lower education
4 – higher good

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$$

$$H_A: not\ (\mu_1 = \mu_2 = \mu_3 = \mu_4)$$

#5 Do medu & fedu affect absence?

ANOVA for absences between fedu

```
Analysis of Variance Table

Response: absences
           Df  Sum Sq Mean Sq F value Pr(>F)
fedu        3    27.7  9.2259  0.4229 0.7366
Residuals 633 13808.8 21.8149
```

ANOVA for absences between medu

```
Analysis of Variance Table

Response: absences
           Df  Sum Sq Mean Sq F value  Pr(>F)
medu        3   184.3  61.423  2.8479 0.03684 *
Residuals 633 13652.2  21.568
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

<u>Conclusion:</u>

Another surprising result. Even though we are confirmed from the mosaic plot that parents' education levels are generally aligned, the ANOVA tests give us an unexpected result:

- P-value of ANOVA for absence between fedu is much higher than the significance level of 0.05, so we do not reject the null hypothesis
- Yet, p-value of ANOVA for absence between medu is less than the significance level of 0.05, so we reject the null hypothesis

One way to explain the result might be: Mothers play a more important role in their children's education

Wrap-up and Overall Conclusion

1. There is a positive correlation between age and absence
2. There is no enough evidence showing that travel time would suggest different absence mean
3. No sufficient evidence to claim that different health status would suggest different absence mean
4. Students with Internet access might have a greater absence mean
5. Father’s education may not have effects on your absence, but mother’s might does

Future & Impression

Impression

The results are not that surprising in general, because there are a considerable number of factors of absence.

I want to

Does number of absence have association with students academic performance?

The factors which can prevent a high number of absence