

Motivation/problem:

Research question/Goal: To explore the relationship between the demand on GP surgeries and other variables.

Dataset intro (brief): The dataset med\_care\_demand.txt contains data on demand on medical care usage by a cohort of approximately 4000 individuals aged 66 or over.

## Background/Literature Review

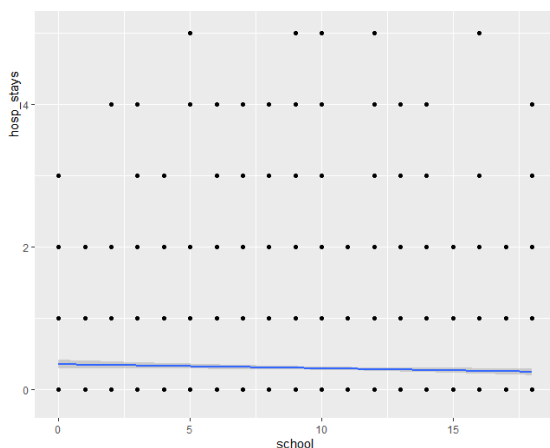
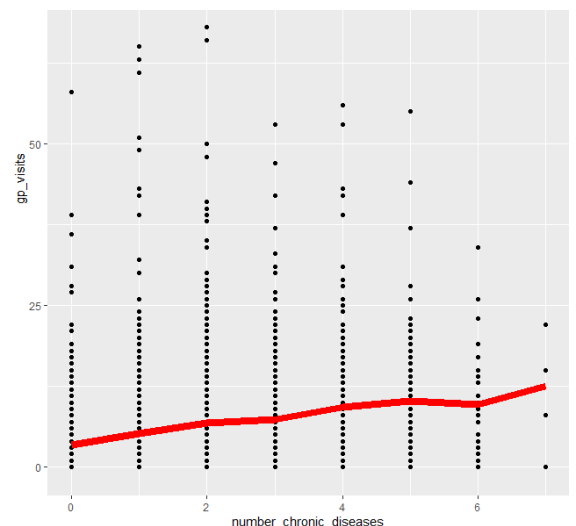
- Data subject: Read about the influencing factors on demand on GP surgeries, which includes but is not limited to population size, age, time of year, education level of local populace, and relationship status. (M. H. BANKS, 1976) (Parkin, 1979) (Beccy Baird, 2016)
- Data methods: Since the outcome is counts, a Poisson model is appropriate. Linear regression will also be useful.
- Performance metrics: No one data method used in isolation can give us a comprehensive understanding of our data. Clustering is good for identifying similar groups, but it does little to explain the causes of these clusters. Similarly, linear regression helps us identify relationships between variables but doesn't identify causality. A mix of data methods and critical analysis will be necessary.

## Results

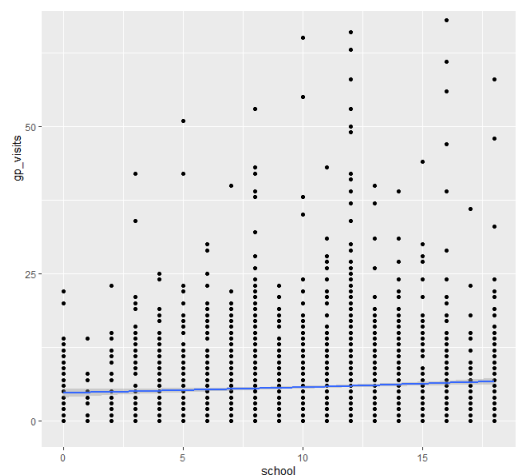
### Relationships between variables

As seen in the graph to the right, there seems to be a slight positive correlation between the number of chronic diseases a patient has and GP visits. This is not entirely surprising.

As seen in the graphs below, patients with more years of school completed had more GP visits but less hospital stays, which supports academic findings that suggest that patients from lower socio-economic backgrounds are less likely to seek medical care, so when they do their condition is likely to be more serious.

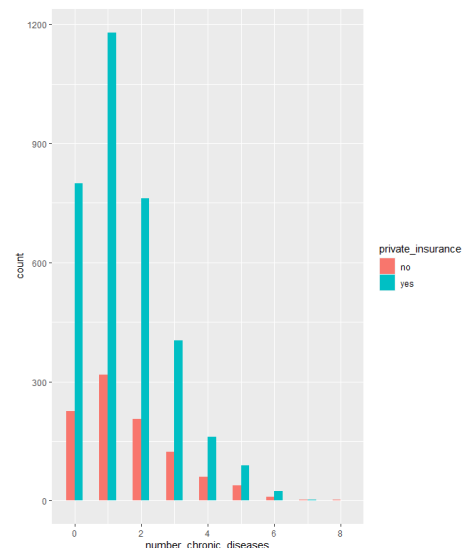
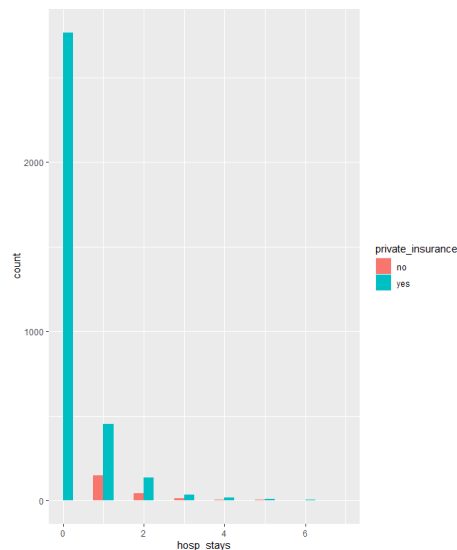
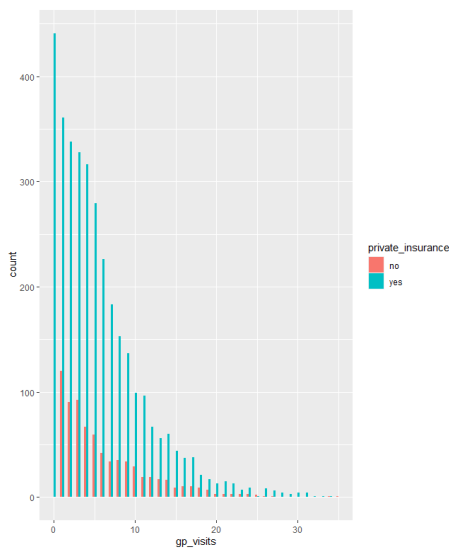
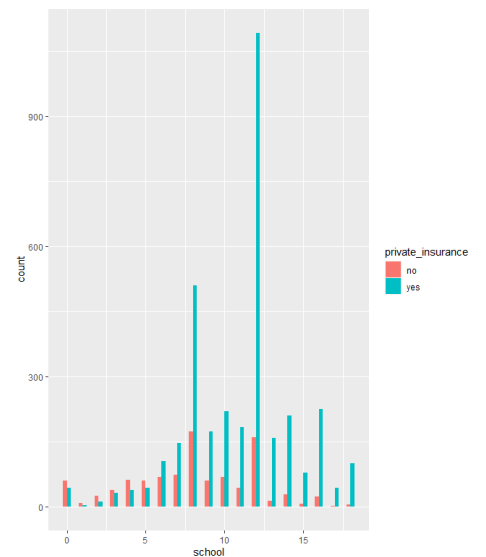


Private Health Insurance  
3.5x as many people have private health insurance than those without. As seen below on the right,



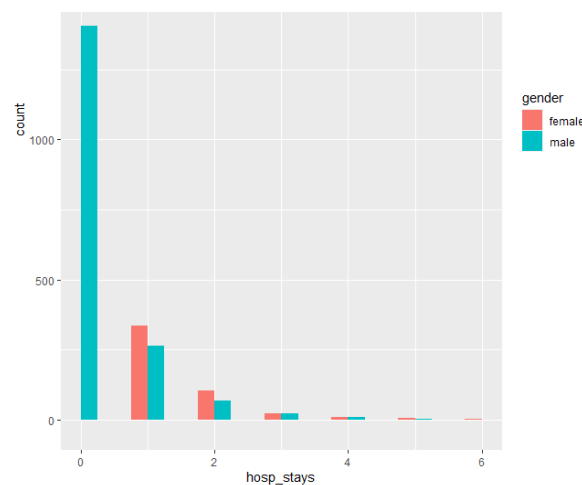
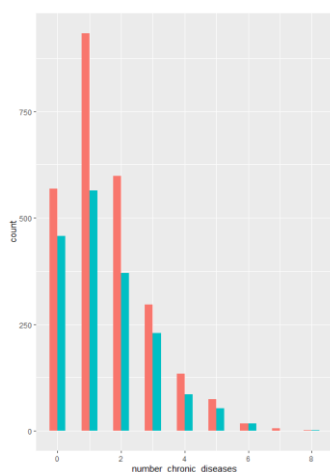
the more years of school a patient has completed, the more likely they are to have private health insurance. This must be taken into consideration when examining the relationship between private health insurance and other variables.

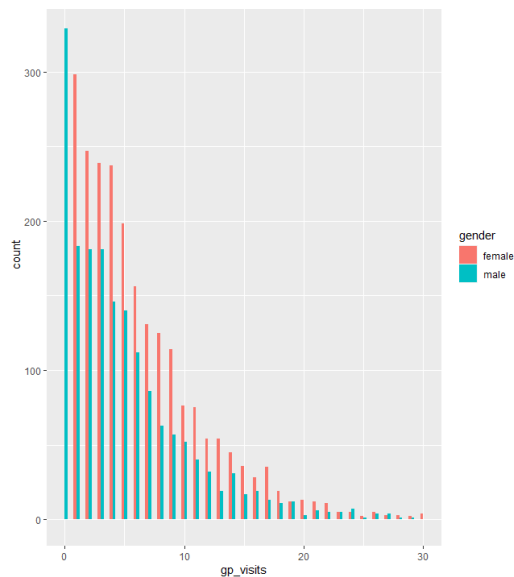
As seen in the graphs below, patients with private health insurance were more likely to have more chronic diseases, but not more likely to visit the GP or stay in hospital.



## Gender

Women were more likely to visit a GP than men, which was expected from our literature review. Men are more likely to have 3+ chronic illnesses and have more stays at hospital. Maybe if men went to the GP more frequently, the amount of chronic diseases and hospital stays they have would go down.





## Discussion

### Challenges

About 35% more women were included in the data than men, which made it difficult to compare the 2 groups. However, the sample size was large enough that this wasn't too much of an issue. The health of a patient was self-reported, which made it a subjective and inaccurate metric which couldn't be used reliably.

### Methods tried and abandoned

Clustering was considered, however no clear cluster was found with any combination of variables. This makes sense because there are no clusters of health. Health is a wide 3D spectrum and people's health cannot be grouped into any number of clusters. Similarly, classification was considered however it would require great effort to build a classification model which was not the purpose of this report, which is simply to analyse the data.

### Conclusion

Men and those from worse socio-economic backgrounds should be encouraged to visit their GP more frequently to reduce the development of chronic diseases and to make more space in hospitals. Private health insurance should be made more readily available to those with lower levels of education.

### Future work

Income, diet, exercise, postcode, and general happiness have large influences on peoples' health, and I hope these datapoints are included so we can analyse the data further. An accurate model could be built to classify users health, however this was not attempted due to time limitations. Measures that could be undertaken to implement the recommendations made in the conclusion should be researched further.