

Capstone Project - Statistical Analysis

Drew Matheson

10/23/2017

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

##
## Attaching package: 'miscset'

## The following object is masked from 'package:dplyr':
##
##   collapse
```

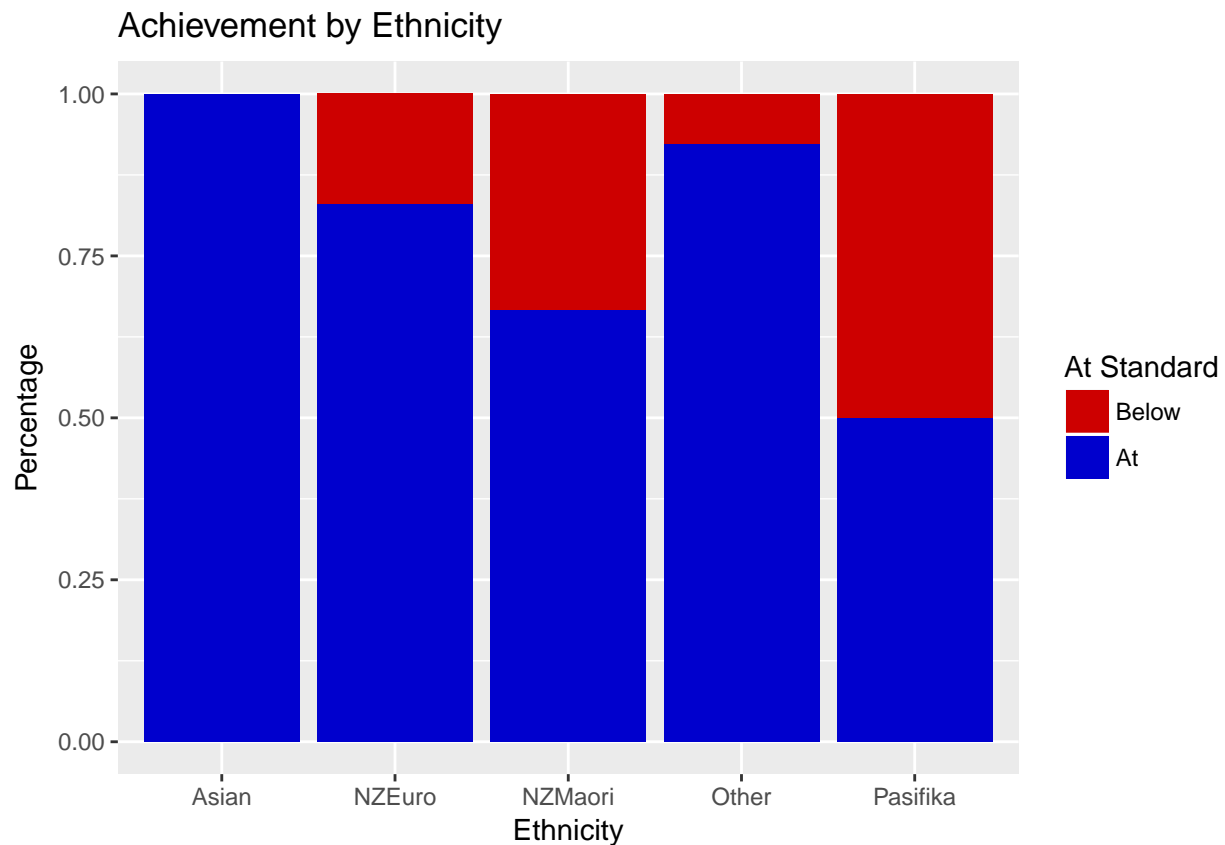
Statistical Analysis

Percentage of students achieving 'At'

The first step was to identify the percentage of students in the dataset who were achieving 'At' or 'Above' the standard. 28 students (19%) were achieving 'Below' or 'Well Below', 119 students (81%) were achieving 'At' or 'Above'.

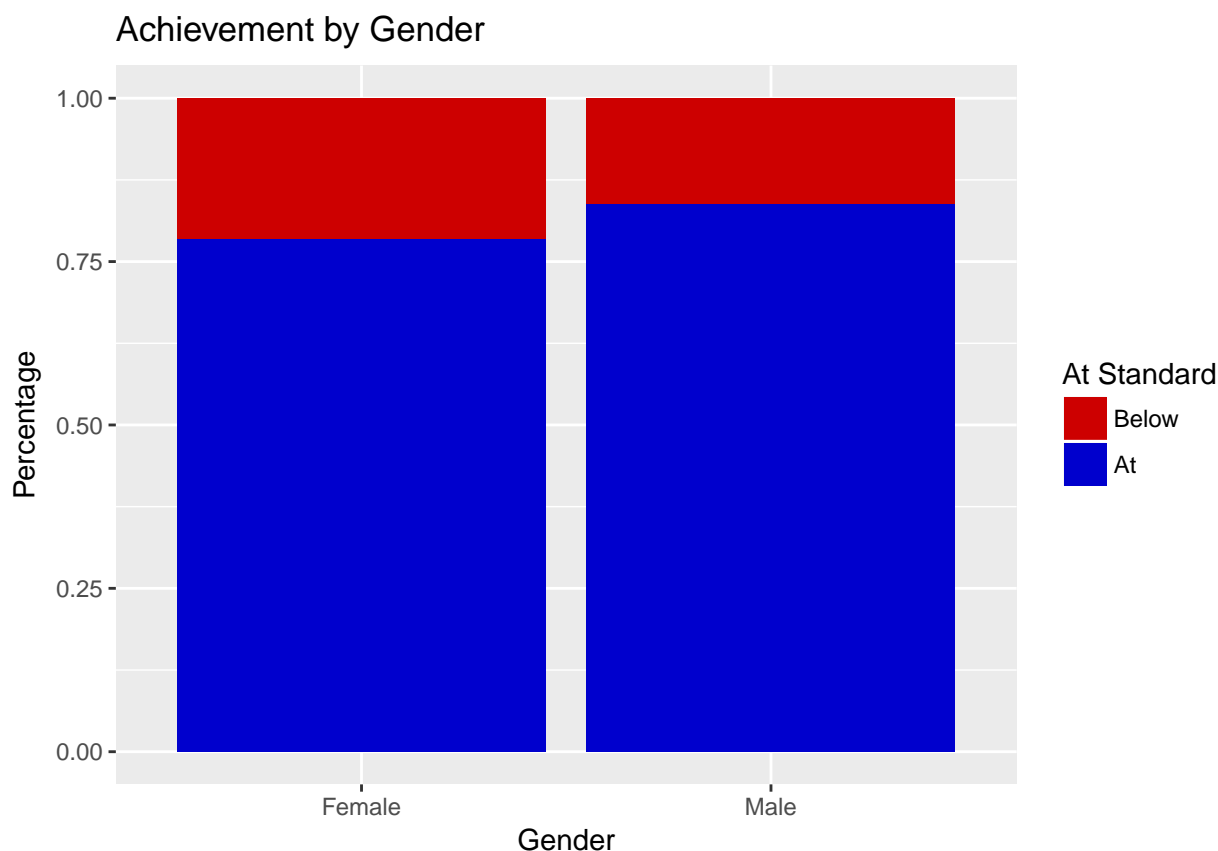
Investigating achievement within ethnic groups

I grouped the data by the 'ethnic' variable and then took the mean of the 'At' variable, as this is a binary variable this gave the percentage of each ethnic group that was achieving the standard. This showed that NZMaori and Pasifika students were more likely to achieve below the standard. 33% of NZMaori students were below the standard, as were 50% of Pasifika students. This was in contrast to NZEuro students of which 17% were achieving below. This suggests that NZMaori and Pasifika students are much more likely to achieve below the standard. I plotted this as follows:



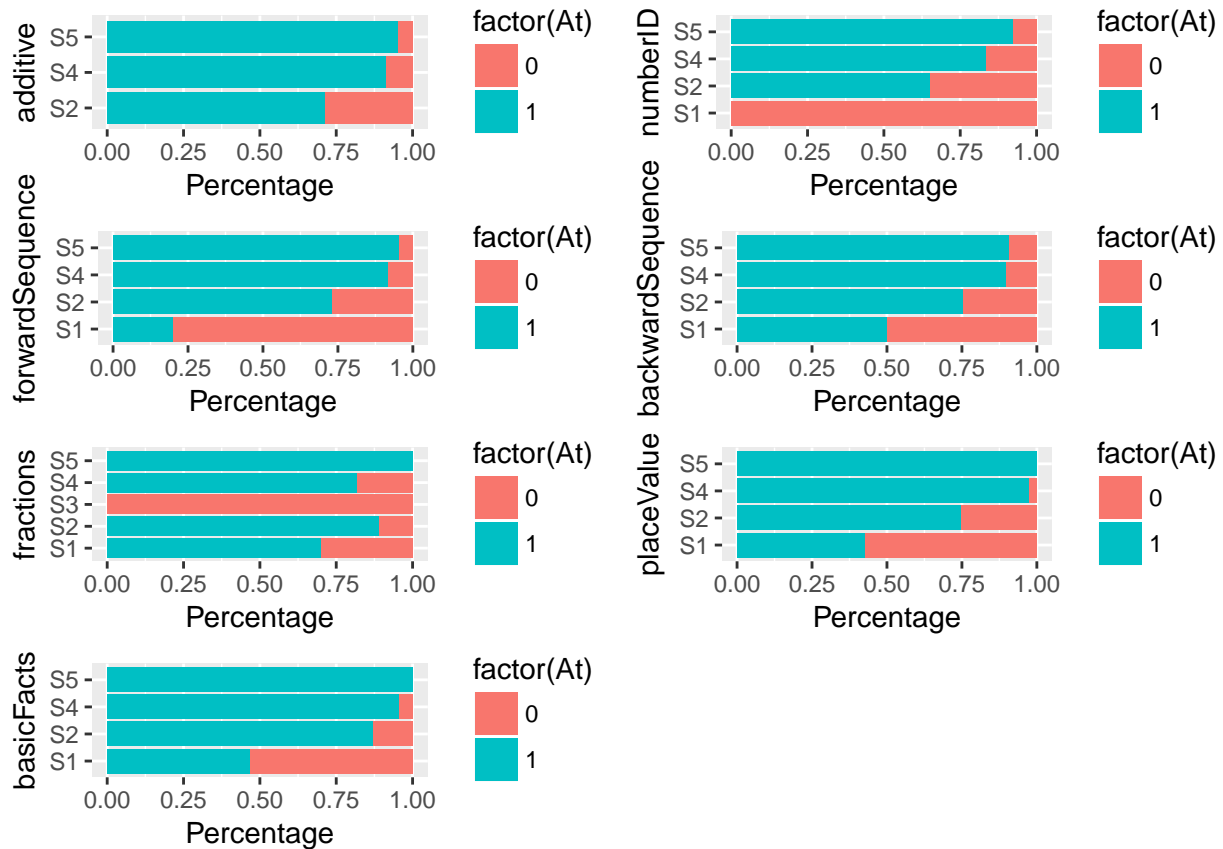
Investigating achievement within gender groups

Data was grouped by the 'gender' variable and then the mean of the 'At' variable was calculated, as this is a binary variable this gave the percentage of each ethnic group that was achieving the standard. This showed that 21% of female students and 16% of male students were achieving below the standard. Female students are slightly to achieve below the standard. I plotted this as follows:



Investigating achievement of students within each JAM assessment domain.

Data was grouped by each of the 7 JAM assessment domain variables and then the mean of the 'At' variable was calculated. This gave the percentage of students at each stage within each domain that was achieving the standard. ###Findings As expected, with nearly every domain from the JAM assessment, the higher stage that a student is assessed as being at the more likely it is that they will be achieving 'At' the standard. The one exception is the 'fractions' domain, there was not a clear trend in this domain. ###Bar Plots To visualise this I plotted each domain in a combined plot. Each individual bar plot showing the percentage of students at each stage of that domain who are judged 'At' or 'Below' the standard.



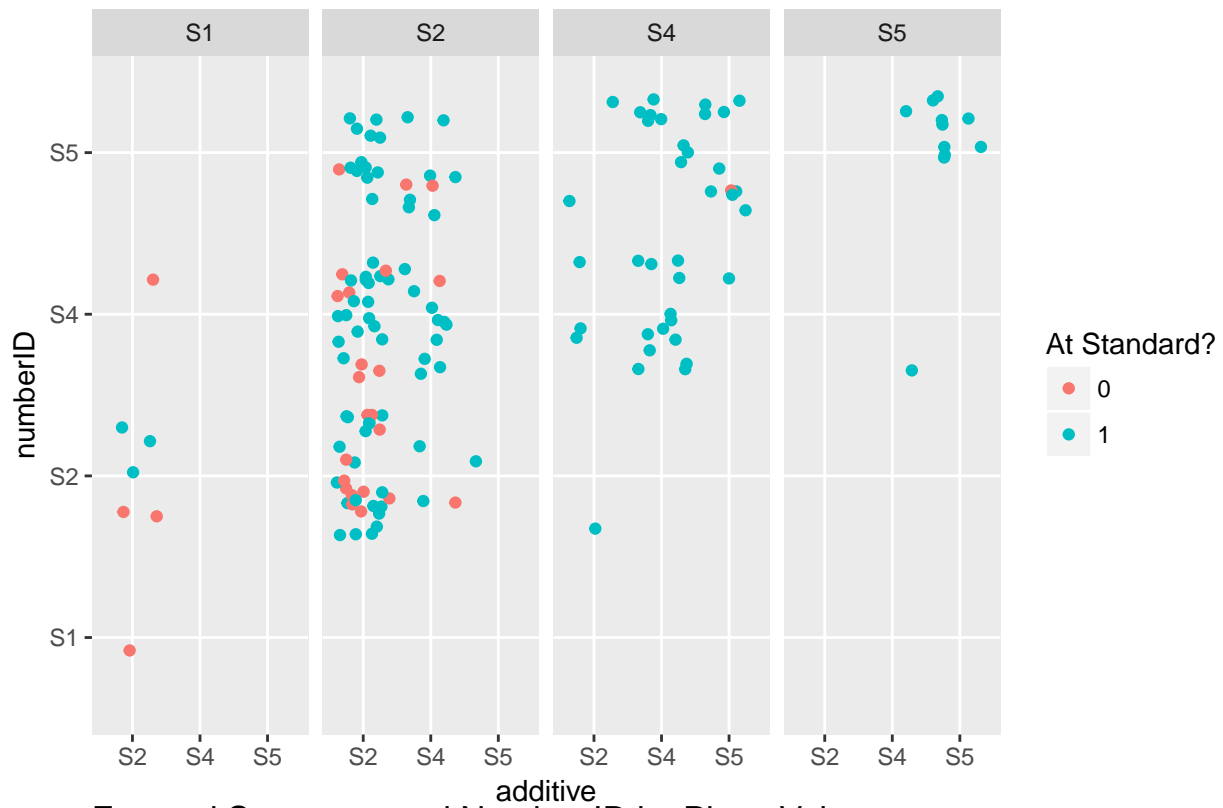
Interpreting the plot

The plots show visually that the higher the stage the more likely that a student will be 'At' the standard. The 'numberID', 'placeValue', 'forwardSequence', and 'basicFacts' domains have the largest proportion of students who were at stage 1 (S1) and then judged as 'Below' the standard. Over half of the students who were Stage 1 in these domains did not reach the standard.

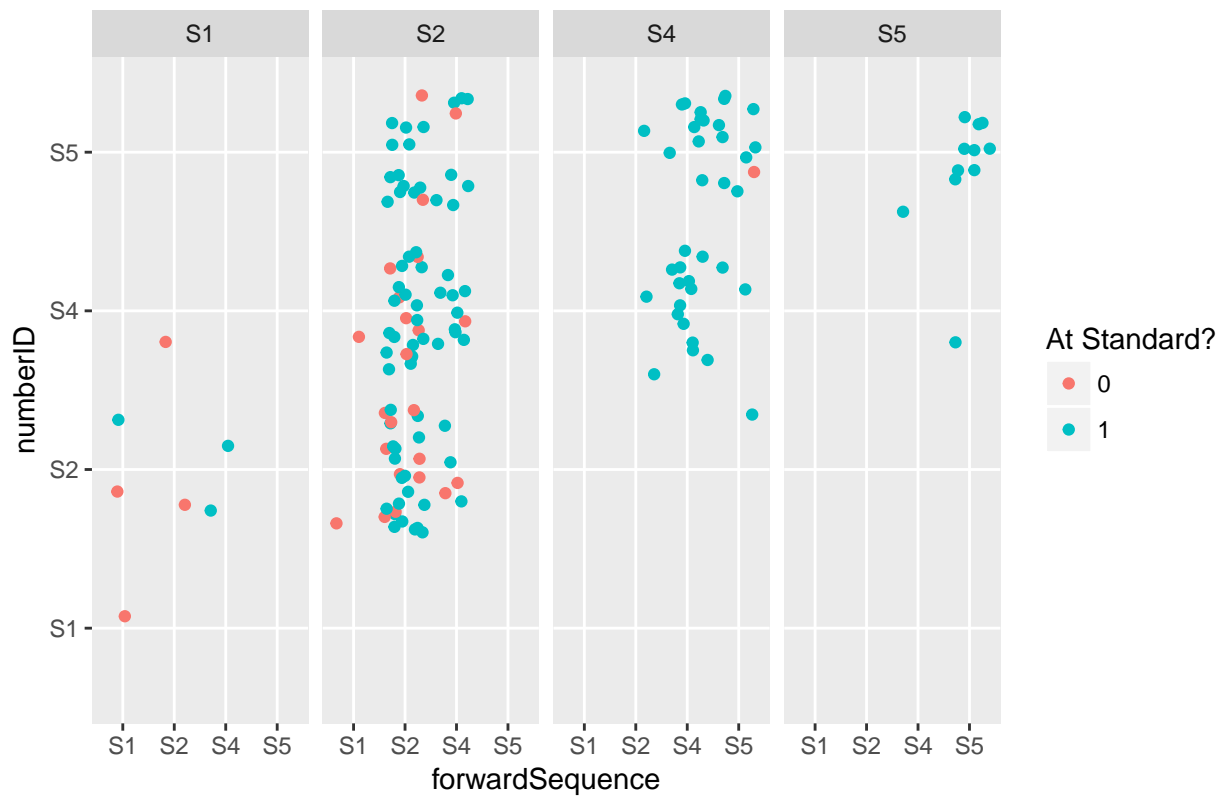
Scatter Plots

I used facet grid scatter plots to visualise three variables and colored the points to show which observations achieved 'At' or 'Below' standard.

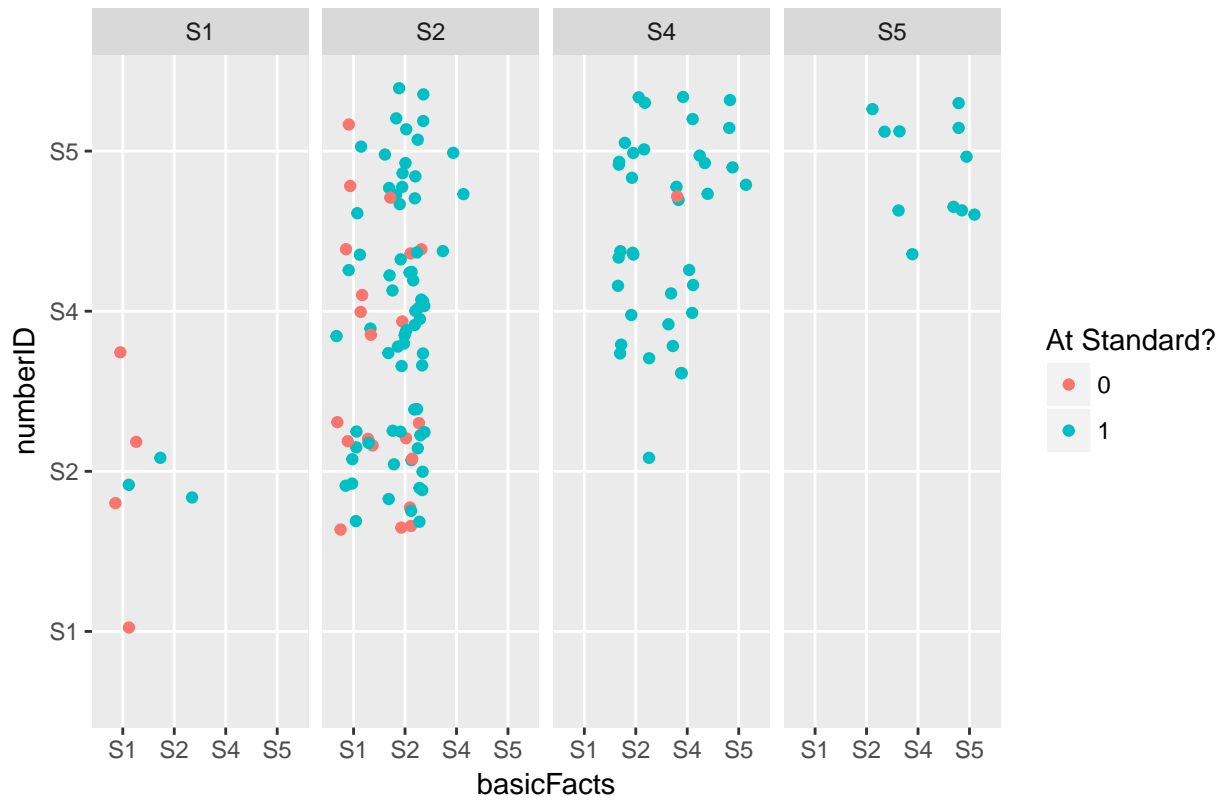
Additive and Number ID



Forward Sequence and Number ID by Place Value



Basic Facts and Number ID by Place Value



Interpreting Scatter Plots

As expected the plots clearly show that the higher the stage in each domain the more likely that a student will achieve 'At' the standard.