

# Blindfire Plotter

*Jimmy Oh*

# Blindfire



Source: <http://forums.3dtotal.com/showthread.php?p=802010>

# What is Blindfire Plotter

*Blindfire Plotter* is a tool that automatically generates a large number of plots - one for every known combination of factors. From a simple pair-wise plot to more complex plots utilising superposition and juxtaposition to cross-classify by multiple factors. Some of these plots will inevitably be useless, but others may provide interesting and unexpected insights into the data.

## Case Study: Secondary Schools Data

NZQA (New Zealand Qualifications Authority) administers NCEA (National Certificates of Educational Achievement):

- ▶ New Zealand's nation-wide secondary school qualification system.
- ▶ 3 levels corresponding to the final 3 years of Secondary School.

NZQA releases various data related to NCEA under **Secondary Schools Statistics**.

# The Data

**Years:** 2004 - 2011

**Subjects:** 33

**Rows:** 12,006

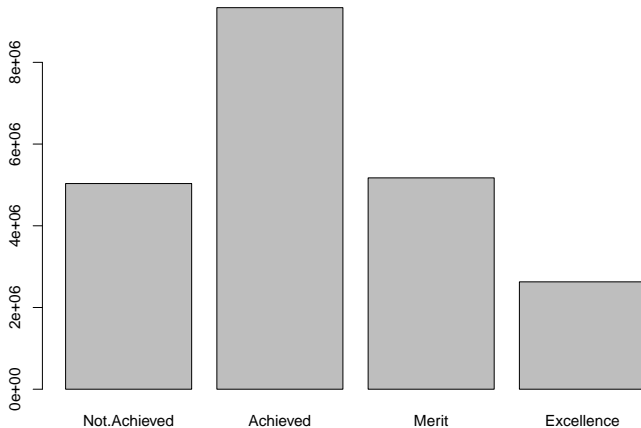
	Year	Subject	Standard	NCEA.Level	Ethnicity
1	2004	Accounting	Ext	Level 1	NZ Maori
2	2004	Accounting	Ext	Level 1	NZ Maori
3	2004	Accounting	Ext	Level 1	NZ European
4	2004	Accounting	Ext	Level 1	NZ European

	Gender	Not.Achieved	Achieved	Merit	Excellence
1	Male	697	365	277	119
2	Female	966	549	322	153
3	Male	3137	3513	3420	2199
4	Female	2465	2720	2622	1967

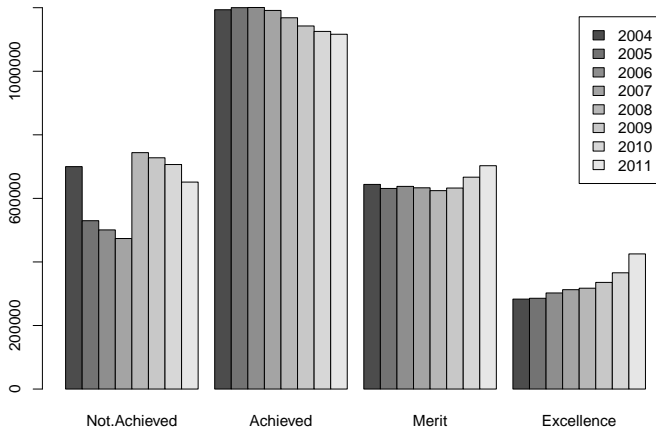
Data are counts of Standards not People.

## Manual Plotting

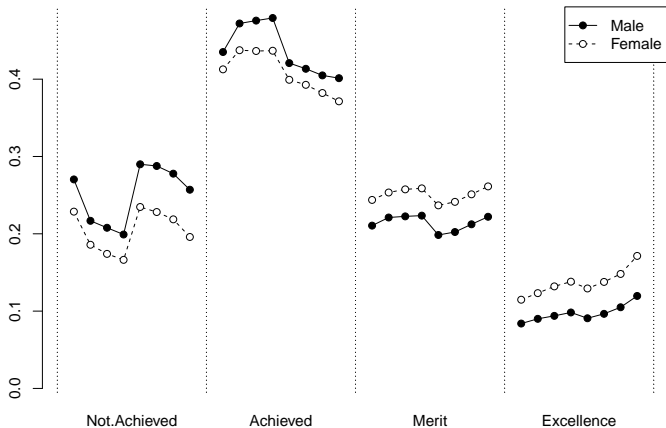
```
> barplot(apply(rep4subj[,c("Not.Achieved",  
    "Achieved", "Merit", "Excellence")], 2, sum))
```



```
> ResultByYear = nzqgrabMulti(rep4subj,  
  list("Not.Achieved", "Achieved", "Merit",  
    "Excellence"), "Year")
```

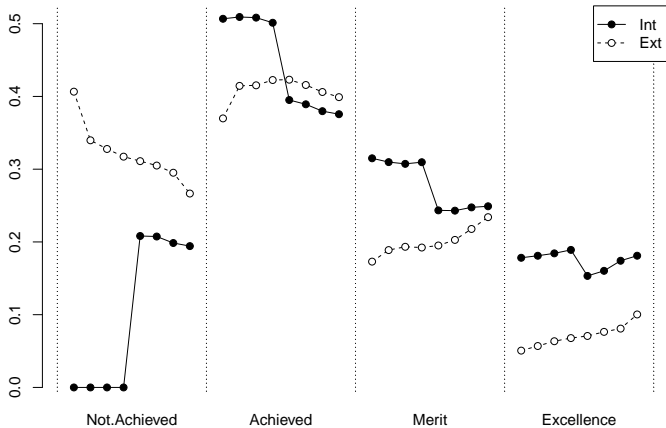


# The Gender Gap





Uh Oh



# Uh Oh

Problem with data discovered several hours in.

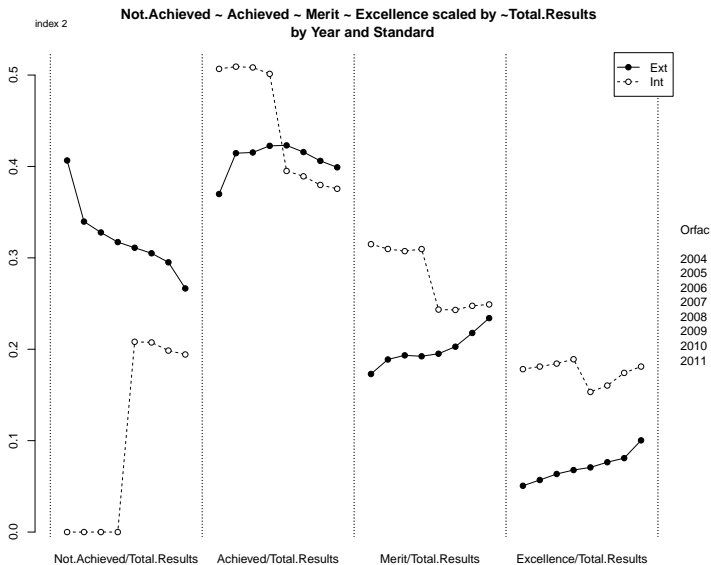
- ▶ Invalidates all previous plots and work.
- ▶ Several ways to correct:
  - ▶ Ignore Internal results.
  - ▶ Ignore the first 4 years (of 8).
  - ▶ Ignore Not Achieved results.
- ▶ First 2 options relatively easy to correct, might be able to salvage some work, but still need to check for errors and lose a lot of information.
- ▶ Last option doesn't remove as much, but considerably more time consuming to alter all the code.

## Uh Oh

To summarise, by using a manual process:

- ▶ Very time consuming.
- ▶ Only discovered a critical problem with the data after having already invested a lot of time - may not have discovered it at all! (Problem only found after cross-classifying Not Achieved with both Standard Type and Year).
- ▶ Time consuming to correct for discovered problem.

# The Blindfire Way



# Simple Arguments

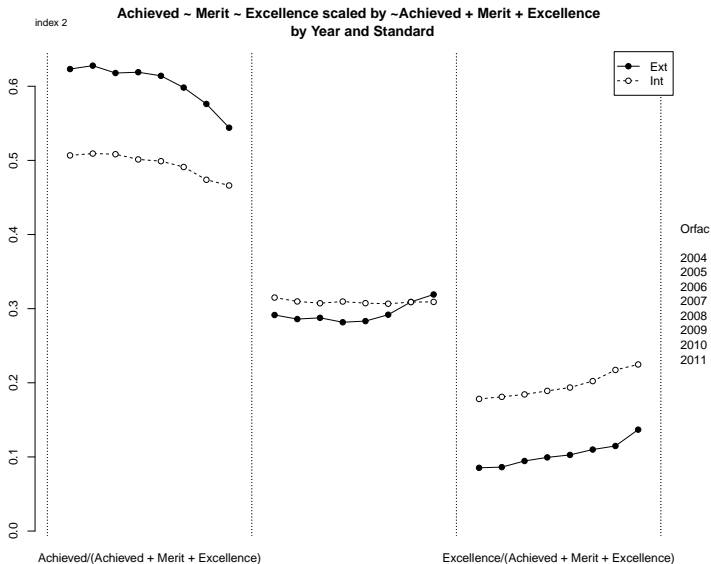
```
respform Not.Achieved ~ Achieved ~ Merit  
        ~ Excellence  
scaleresp ~ Total.Results  
        orfac "Year"  
        unorfac "Standard"
```

Produces 2 plots in about 0.1 seconds.

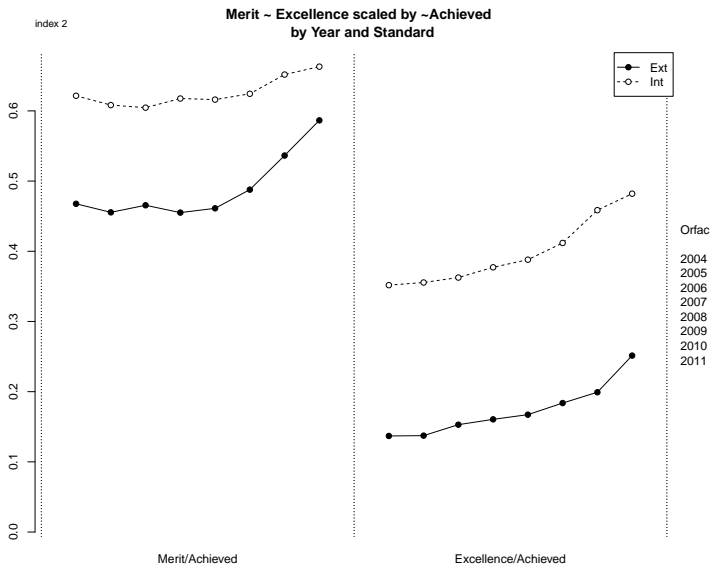
## Can adjust arguments easily

```
respform Achieved ~ Merit ~ Excellence  
scaleresp ~ Achieved + Merit + Excellence  
  orfac "Year"  
  unorfac "Standard"
```

# Can adjust arguments easily

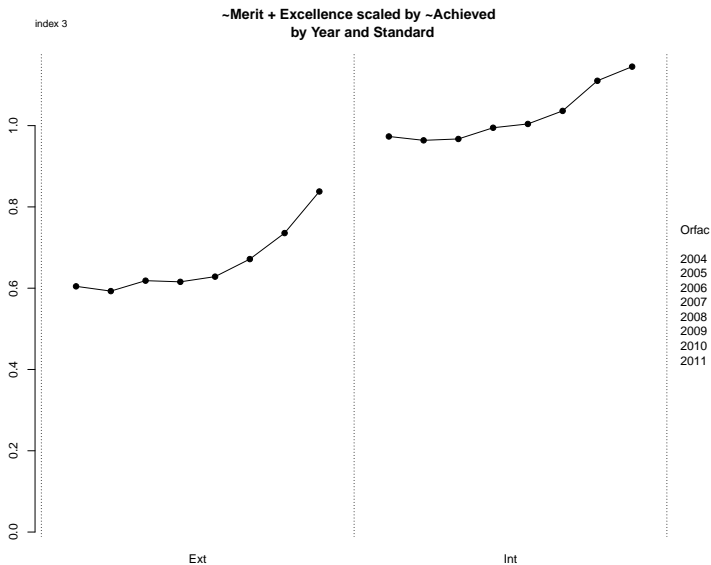


# Another way to look at it



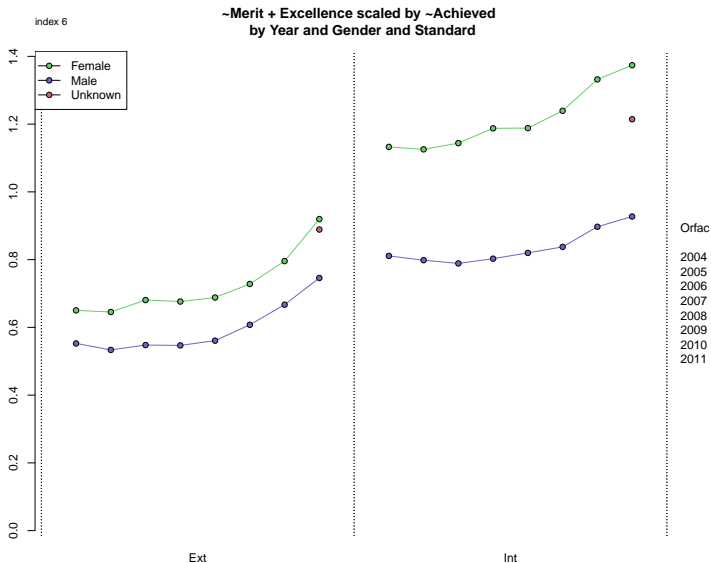


# Yet another way to look at it



# The Gender Gap (Again)

unorfac = c("Standard", "Gender")

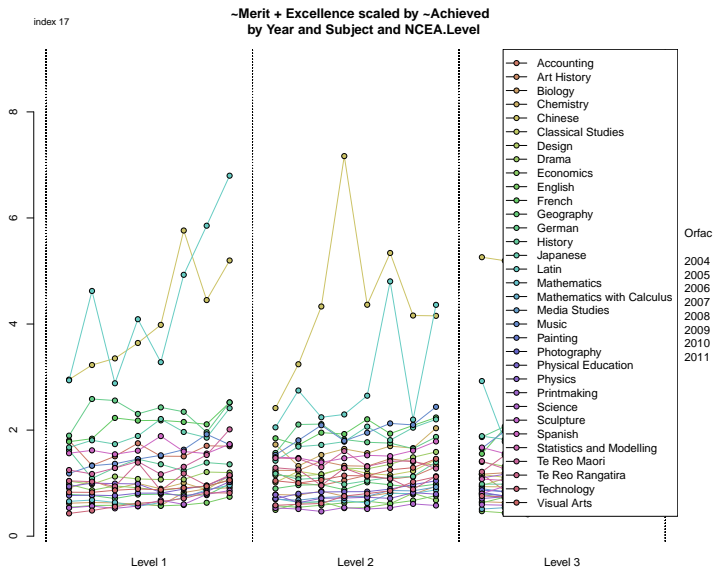


## A Full Call

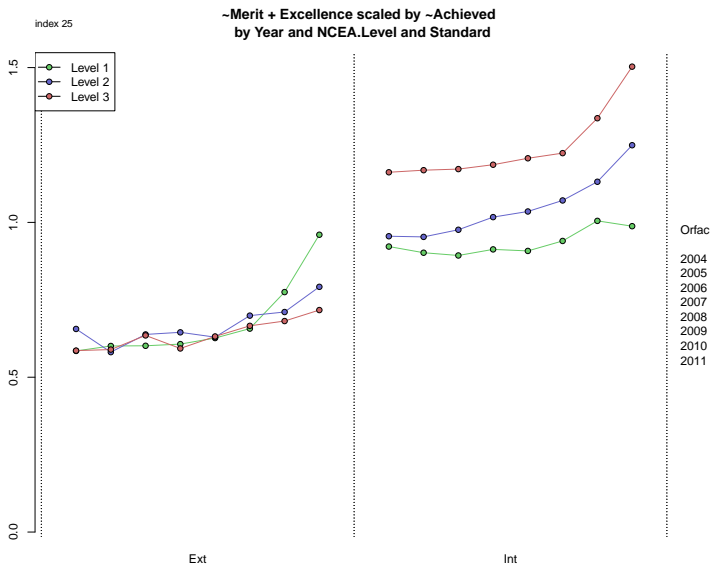
```
respform ~ Merit + Excellence  
scaleresp ~ Achieved  
  orfac c("Year", "NCEA.Level")  
  unorfac c("Subject", "Standard",  
            "Ethnicity", "Gender")
```

Produces 56 plots in about 4 seconds.

# Some are useless...



# Others may be interesting



# The Blindfire Way

To summarise, by using *Blindfire Plotter*:

- ▶ Plotting and re-plotting becomes effortless.
- ▶ Very easy to manipulate data for different 'views'.
- ▶ We can focus on looking at plots - rather than on making them.

# A Paradigm Shift

Focus shifts from:

- ▶ A single plot that retains a lot of information, e.g. Dotcharts

To a package of plots:

- ▶ **Screening Plots** - Simple plots to quickly identify whether there's anything interesting.
- ▶ Once identified, use a variety of more complex plots to examine.
- ▶ Could utilise interactivity to enable users to 'zoom' in on a screening plot to get a more complex plot.

## Concluding Remarks

- ▶ *Blindfire Plotter* can lead to significant savings in time.
- ▶ Because it plots every possible combination, can be used to identify and examine problems that may otherwise have been missed.
- ▶ This is just a brief look, haven't covered features like modular extension, customising plotting functions or using *Blindfire Plotter* as a way to easily reformat data.
- ▶ Still in development - this is just a glimpse of the potential, e.g. speed optimisation, interactivity.



## References

- ▶ R Development Core Team (2012). **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- ▶ Secondary School Statistics, **NZQA**, URL <http://www.nzqa.govt.nz/>