R Data Example

```r
library(readstata13)
library(xtable)
library(plyr)
library(dplyr)
```

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

## The following objects are masked from 'package:plyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize

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

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

```r
library(ggplot2)
library(data.table)
```

```
##
## Attaching package: 'data.table'

## The following objects are masked from 'package:dplyr':
##
##     between, first, last
```

```r
savepath = "/home/ms486/Dropbox/papers/education/EducSearch/Tex/educ28/figures_paper/"
datapath = "../Data/descriptive_data/"
savetab  = "/home/ms486/Dropbox/papers/education/EducSearch/Tex/educ28/tables_paper/"

mytheme = theme(legend.title=element_blank(),
                legend.position="bottom",
                panel.border = element_blank(),
                panel.grid.major = element_blank(),
                panel.grid.minor = element_blank(),
                legend.text = element_text(size=12,face="bold"),
                strip.text = element_text(size = 12,face="bold"))


dat <- read.dta13("~/Dropbox/papers/education/EducSearch/Data/nlsy_new/processdata/nlsy_merged.dta")
```

```
## Warning in read.dta13("~/Dropbox/papers/education/EducSearch/Data/nlsy_new/processdata/nlsy_merged.d-
##   job_stat:
##   Missing factor labels - no labels assigned.
##   Set option generate.factors=T to generate labels.
```

```r
names(dat)
```

```
##  [1] "caseid"        "year"          "weekyear"      "week"
##  [5] "joball_hours"  "job_stat"      "job_wave"      "job_num"
##  [9] "job_occ_raw"   "occ70"         "occ00"         "job_occ"
## [13] "job_req_educ"  "job_hours"     "job_wage"      "job_prev_num"
```

```
## [17] "job_cps"        "int_week"       "last_week"       "mil_curr"
## [21] "mil_first"       "mil_ever"       "job_change"      "job_seq"
## [25] "ind_sample"      "ind_race"       "ind_sex"         "ind_dob_year"
## [29] "ind_dob_month"   "ind_higrade_rc" "ind_higrade"     "ind_ed_rc_yr"
## [33] "ind_ed_yr"       "ind_ed_yr_ret"  "region_rc"       "region"
## [37] "ind_ed"          "ind_ed_rc"      "ability"         "month"
## [41] "employed"        "J2J"            "U2E"             "E2U"
## [45] "finished_ed"     "firstweek"      "week0"           "weekyear0"
## [49] "year0"           "monthyear0"     "month0"          "transition"
## [53] "spell"           "spell_dur"      "u_dur"           "j_dur"
## [57] "spdur"           "job_lwage"      "match_over_ed"   "match_same_ed"
## [61] "match_under_ed"
```

```r
dat1 = subset(dat,select=c(year,job_req_educ),!is.na(job_req_educ))
tab  =  dat1 %>% group_by(year,job_req_educ) %>% summarise(n = n()) %>% mutate(freq = n / sum(n))
tab
```

```
## # A tibble: 116 x 4
## # Groups:   year [30]
##     year job_req_educ                          n     freq
##    <int> <fct>                             <int>    <dbl>
##  1  1977 Less than a High School Diploma      46 0.195
##  2  1977 High School Diploma                 188 0.797
##  3  1977 Bachelor's Degree                     2 0.00847
##  4  1978 Less than a High School Diploma    3727 0.174
##  5  1978 High School Diploma               17428 0.812
##  6  1978 Bachelor's Degree                   302 0.0141
##  7  1979 Less than a High School Diploma    5821 0.179
##  8  1979 High School Diploma               25938 0.796
##  9  1979 Bachelor's Degree                   782 0.0240
## 10  1979 Advanced Degree                      51 0.00156
## # ... with 106 more rows
```

```r
graph = ggplot(tab, aes(x=year, y=freq,color=as.factor(job_req_educ))) + geom_line(size=2)  + ylab("Rat
graph = graph +  mytheme + theme(axis.title = element_text(size = rel(1.1)),
                                 axis.text = element_text(size = rel(1.1)),
                                 strip.text = element_text(size = 12),
                                 panel.spacing = unit(1, "lines"))
graph
```

```
## Warning: Removed 32 rows containing missing values (geom_path).
```
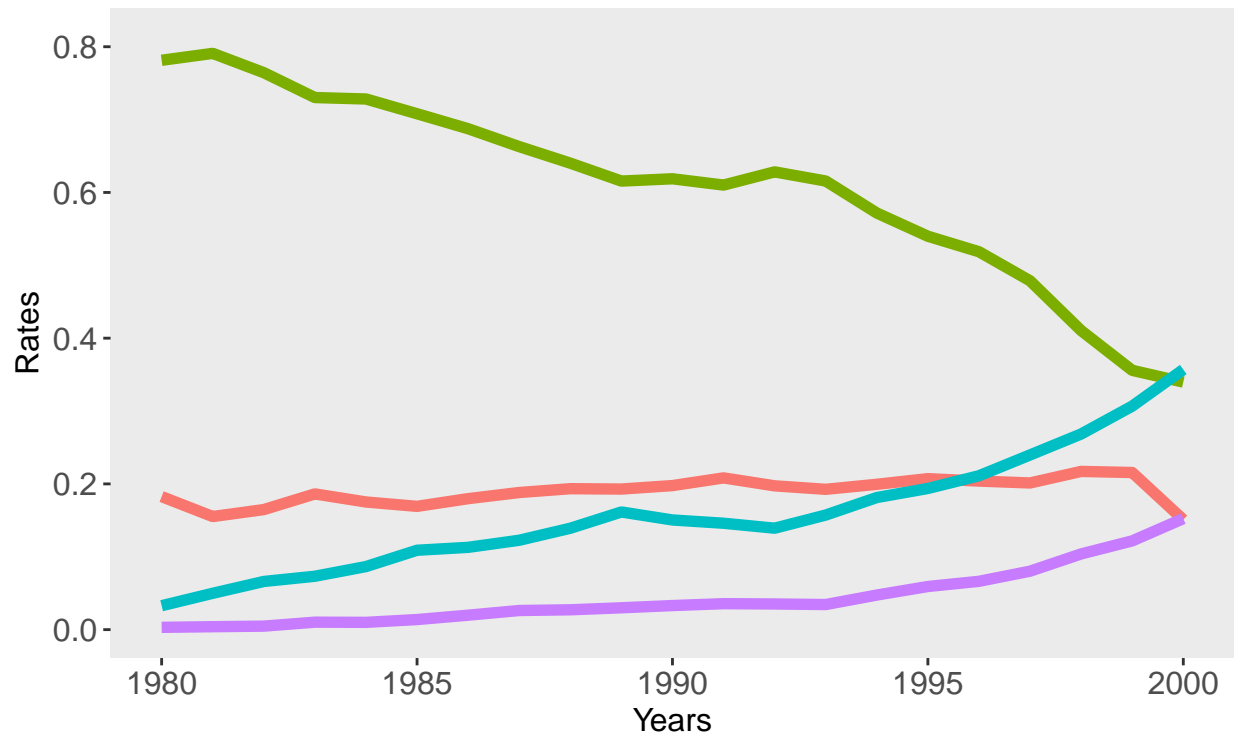
```
## education
tab        = read.csv("~/Dropbox/Git/tab.csv", header=FALSE)
tab        = cbind(seq(2010,1980,by=-1),tab)
names(tab) = c("Years","Less than a High School Diploma", "High School Diploma", "Some College", "Bachel
tab
```

```
##    Years Less than a High School Diploma High School Diploma Some College
## 1   2010                      0.1286019           0.3123940    0.2596948
## 2   2009                      0.1332173           0.3107951    0.2605845
## 3   2008                      0.1341840           0.3116732    0.2597692
## 4   2007                      0.1427711           0.3164401    0.2534145
## 5   2006                      0.1453787           0.3173672    0.2572947
## 6   2005                      0.1479508           0.3215608    0.2538774
## 7   2004                      0.1484674           0.3200572    0.2545592
## 8   2003                      0.1544356           0.3201788    0.2533156
## 9   2002                      0.1589309           0.3209364    0.2527808
## 10  2001                      0.1585859           0.3230370    0.2565636
## 11  2000                      0.1589568           0.3314843    0.2536381
## 12  1999                      0.1659818           0.3334312    0.2484892
## 13  1998                      0.1716383           0.3378065    0.2468251
## 14  1997                      0.1789366           0.3375894    0.2448939
## 15  1996                      0.1825300           0.3360147    0.2457893
## 16  1995                      0.1833295           0.3391653    0.2478340
## 17  1994                      0.1911095           0.3435291    0.2432270
## 18  1993                      0.1977227           0.3536886    0.2300091
## 19  1992                      0.2058734           0.3597655    0.2208584
## 20  1991                      0.2156828           0.3860967    0.1838106
## 21  1990                      0.2239215           0.3840562    0.1793506
```

```
## 22   1989                          0.2312009            0.3849088       0.1726433
## 23   1988                          0.2381411            0.3886914       0.1701366
## 24   1987                          0.2437795            0.3866692       0.1708360
## 25   1986                          0.2527182            0.3842817       0.1686766
## 26   1985                          0.2609023            0.3822749       0.1630726
## 27   1984                          0.2668850            0.3840657       0.1582559
## 28   1983                          0.2790465            0.3771917       0.1559991
## 29   1982                          0.2904040            0.3794577       0.1526803
## 30   1981                          0.3029918            0.3755888       0.1508074
## 31   1980                          0.3136464            0.3675695       0.1486028
##      Bachelor's Degree
## 1            0.2993092
## 2            0.2954031
## 3            0.2943736
## 4            0.2873743
## 5            0.2799594
## 6            0.2766110
## 7            0.2769162
## 8            0.2720699
## 9            0.2673518
## 10           0.2618134
## 11           0.2559208
## 12           0.2520978
## 13           0.2437301
## 14           0.2385801
## 15           0.2356660
## 16           0.2296711
## 17           0.2221344
## 18           0.2185796
## 19           0.2135027
## 20           0.2144099
## 21           0.2126718
## 22           0.2112470
## 23           0.2030309
## 24           0.1987153
## 25           0.1943236
## 26           0.1937502
## 27           0.1907934
## 28           0.1877626
## 29           0.1774580
## 30           0.1706120
## 31           0.1701813
```

```r
tab2        = cbind(tab[,1:2],tab[,3]+tab[,4],tab[,5])
names(tab2) = c("Years","Less than a High School Diploma", "High School Diploma", "Bachelor's Degree")

tab1     = melt(tab2,id="Years")
```

```
## Warning in melt(tab2, id = "Years"): The melt generic in data.table has been
## passed a data.frame and will attempt to redirect to the relevant reshape2
## method; please note that reshape2 is deprecated, and this redirection is now
## deprecated as well. To continue using melt methods from reshape2 while both
## libraries are attached, e.g. melt.list, you can prepend the namespace like
## reshape2::melt(tab2). In the next version, this warning will become an error.
```
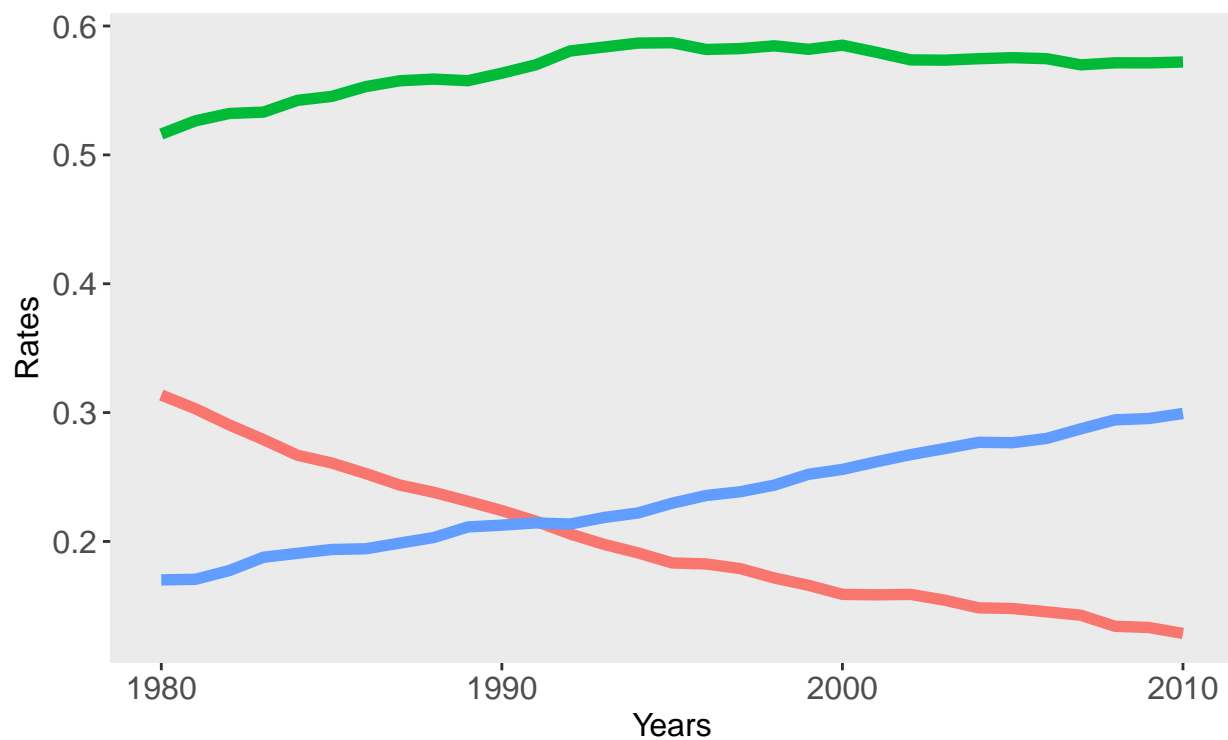
```
mytheme = theme(legend.title=element_blank(),
                legend.position="bottom",
                panel.border = element_blank(),
                panel.grid.major = element_blank(),
                panel.grid.minor = element_blank())


graph = ggplot(tab1, aes(x=Years, y=value,color=variable)) + geom_line(size=2) + ylab("Rates") + xlab("
graph = graph +  mytheme + theme(axis.title = element_text(size = rel(1.1)),
                                 axis.text = element_text(size = rel(1.1)),
                                 strip.text = element_text(size = 12),
                                 panel.spacing = unit(1, "lines"))
graph
```



```
ggsave(paste0(savepath, 'education2.pdf'),graph)
```

```
## Saving 6.5 x 4.5 in image
```