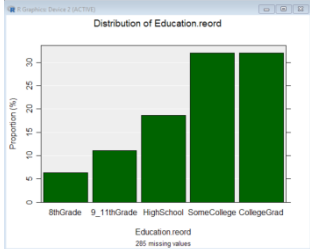# 3.5 Exercise: Relationships between categorical variables – *R version*

This exercise will enable you to construct graphs of two categorical variables as discussed in the previous video. The skills addressed are:

1. Creating plots of two categorical variables (when the predictor variable has only 2 groups).
2. Making a *side-by-side* bar charts and *separate* bar charts for two categorical variables.
3. Filtering out unwanted groups

We will use the **nhanes_1000** dataset from the **FutureLearnData** package.

| #R Code	Output and/or Commentary
# Setup library(iNZightPlots) library(FutureLearnData) data(**nhanes_1000**)	
# Recreate Education.reord nhanes_1000$**Education.reord** = **factor(**nhanes_1000$**Education**, **levels =** c("8thGrade","9_11thGrade","HighSchool","SomeCollege","CollegeGrad")) iNZightPlot(**Education.reord**, data=nhanes_1000)	*Education with levels in a sensible order* *(This was done in Exercise 2.5)*

Education.reord by Gender (side by side barcharts)
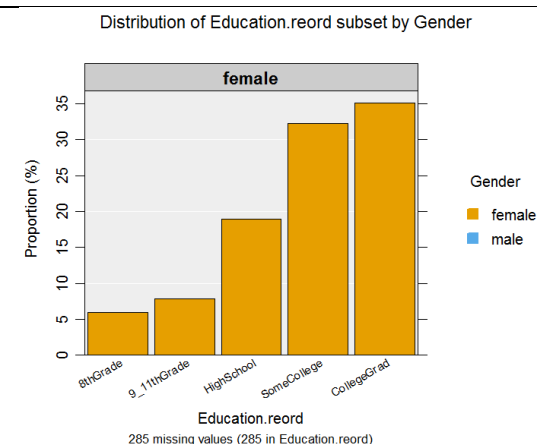
iNZightPlot(**Education.reord, Gender**, data=nhanes_1000)
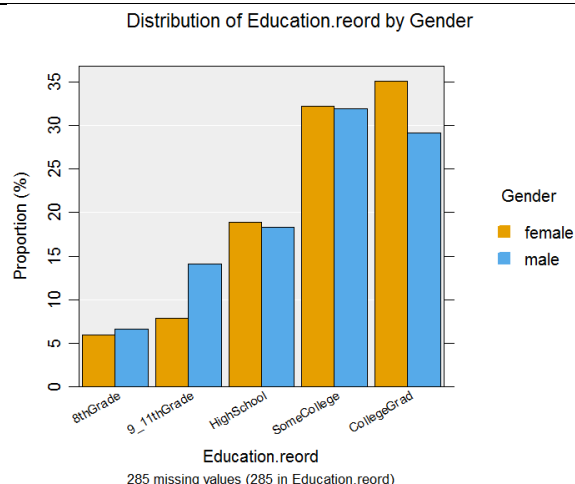
Education **subset by** *Gender* *(also coloured by Gender)*

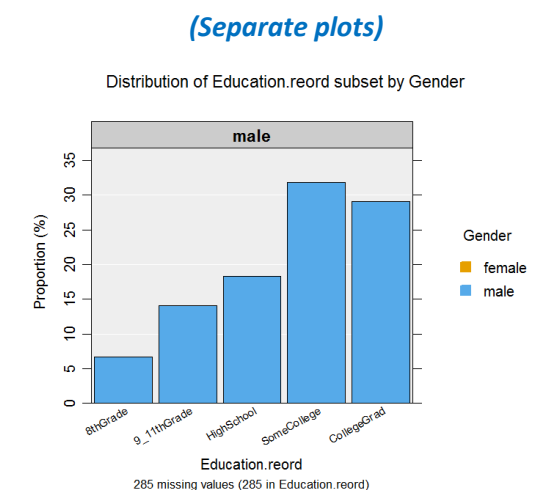iNZightPlot(**Education.reord, g1=Gender**, data=nhanes_1000,
    colby=Gender)
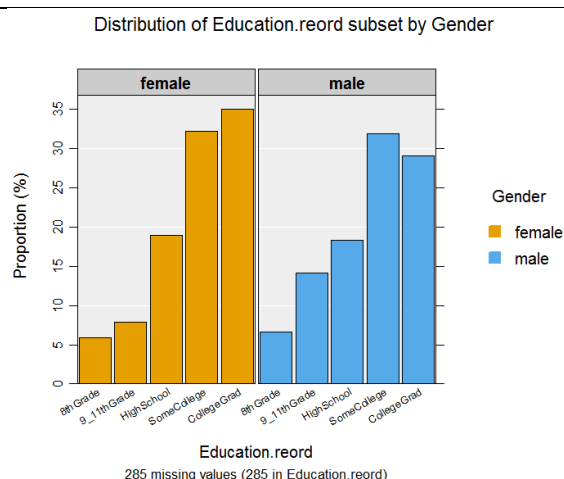
Display just the female plot

iNZightPlot(Education.reord, **g1=Gender, g1.level="female"**,
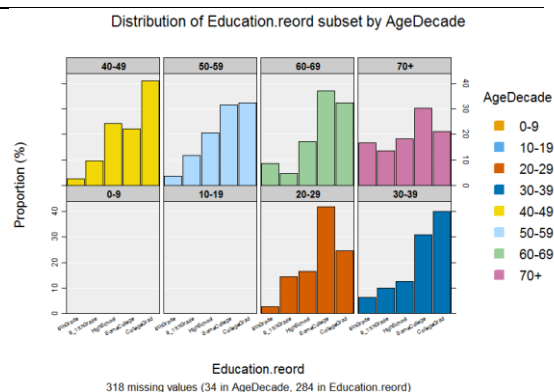   data=nhanes_1000**,** colby=Gender)

(Separate plots)

Display just the male plot

iNZightPlot(Education.reord, **g1=Gender, g1.level="male"**,
    data=nhanes_1000**,** colby=Gender**)**

Plot Education.reord subset by AgeDecade

iNZightPlot(Education.reord, **g1=AgeDecade**, data=nhanes_1000, colby=AgeDecade)

Filter out unhelpful data (Under 20s)

Make a subset of the data without "under 20s" & call it Temp

*(This is an **example of "filtering"** the data)*

Temp = subset(nhanes_1000, AgeDecade!=" 0-9" & AgeDecade !=" 10-19"**)**   # *"!="* is read as *"is not equal to"*

table(Temp$AgeDecade)

This will remove the empty levels
Temp$AgeDecade=factor(Temp$AgeDecade)

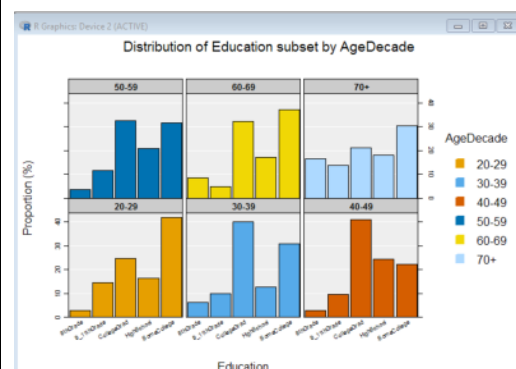# *Replot the data using* the subset of the data called *Temp*
iNZightPlot(Education.reord, g1=AgeDecade, **data=Temp**, colby=AgeDecade)

*(**Warning:** there is a leading space on all of AgeDecade's level names– a data bug)*

Output

0-9	10-19	20-29	30-39	40-49	50-59	60-69	70+
0	0	146	110	144	111	05	66

It still has the empty levels

- **Now construct similar graphics using other categorical variables whose behaviour you may be interested in**

==Optional: *Try this new feature* (interactive web graphics)==

We will export an iNZightPlot graph as an *Interactive HTML* file and open this file up in our default browser. If that is a modern browser like Chrome, Firefox or Safari (but not Internet Explorer) this will then give you an interactive version of the graph that lets you query it in various ways like hovering over bars or clicking them. Explore!

You can give such files to others. They do not need to be connected to iNZight to work.

Here is sample code: **This only works with single graphs, not panels containing multiple graphs**

```
# Make a plot and also store the output in myplot
myplot = iNZightPlot(Education.reord, Gender, data=nhanes_1000)

# Specify a location to store an Interactive HTML file. I will call my file "myintplot.html"
# You will have to change the path to the file because this one is to a location on my desktop!
filepath = "C:/Users/cwil119/Desktop/myintplot.html"

exportHTML(myplot, filepath)
browseURL(filepath)   #open the file up in my default browser
```

To discuss issues related to this Exercise,

go to https://gitter.im/iNZightVIT/d2i-R-discussion

*To be able to post to the list you will have to set up a (free) account on **Github***
https://github.com/login

If your question relates to an Exercise, say which one you are talking about!