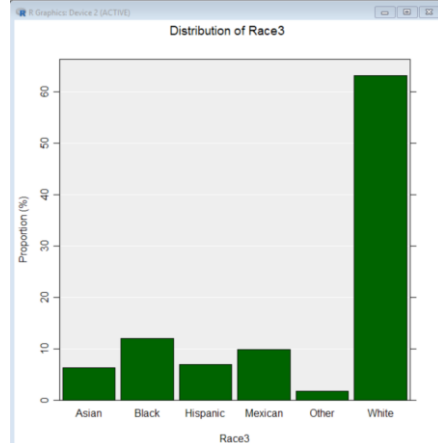
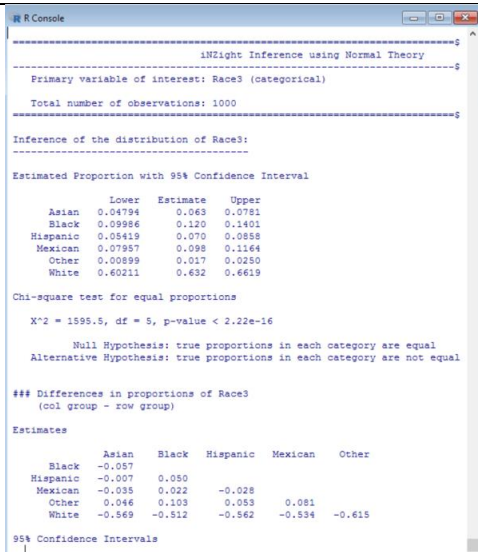
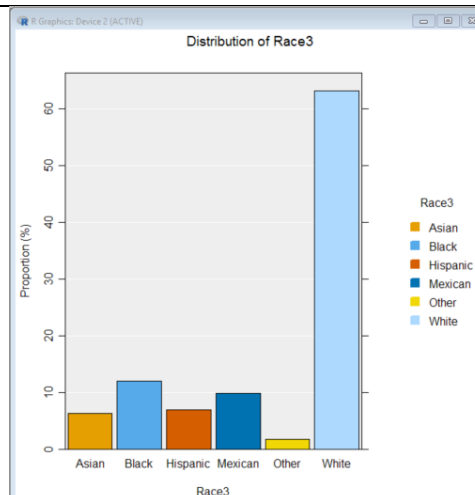


2.5 Exercise: Categorical variables – R version

# R code	Output
<pre># Setup library(iNZightPlots) library(FutureLearnData) data(nhanes_1000)</pre>	
<pre># Plot the variable Race3 # Because Race3 is <i>categorical</i> we get a <i>bar chart</i> iNZightPlot(Race3, data=nhanes_1000)</pre>	
<pre># Get a summary for a variable (Race3) # Equiv. of Get Summary in iNZight getPlotSummary(Race3, data=nhanes_1000)</pre>	<pre>===== iNZight Summary -----\$ Primary variable of interest: Race3 (categorical) Total number of observations: 1000 ===== Summary of the distribution of Race3: ===== Asian Black Hispanic Mexican Other White Total Count 63 120 70 98 17 632 1000 Percent 6.3% 12.0% 7.0% 9.8% 1.7% 63.2% 100% =====</pre>
<pre># Equivalent of Get Inference in iNZight getPlotSummary(Race3, data=nhanes_1000, summary.type="inference", inference.type="conf")</pre>	 <pre>===== iNZight Inference using Normal Theory -----\$ Primary variable of interest: Race3 (categorical) Total number of observations: 1000 ===== Inference of the distribution of Race3: ===== Estimated Proportion with 95% Confidence Interval Lower Estimate Upper Asian 0.04794 0.063 0.0781 Black 0.09986 0.120 0.1401 Hispanic 0.05419 0.070 0.0858 Mexican 0.07957 0.098 0.1164 Other 0.00899 0.017 0.0250 White 0.60211 0.632 0.6619 Chi-square test for equal proportions X^2 = 1595.5, df = 5, p-value < 2.22e-16 Null Hypothesis: true proportions in each category are equal Alternative Hypothesis: true proportions in each category are not equal ### Differences in proportions of Race3 (col group - row group) Estimates Asian Black Hispanic Mexican Other Black -0.057 0.000 -0.028 -0.028 -0.057 Hispanic -0.007 0.050 -0.028 -0.028 -0.057 Mexican -0.035 0.022 -0.028 -0.028 -0.057 Other 0.046 0.103 -0.028 -0.028 -0.057 White -0.569 -0.512 -0.562 -0.534 -0.615 95% Confidence Intervals </pre>

Colour by a variable (*Race3*) (default colour palette)

```
iNZightPlot(Race3, data=nhanes_1000, colby=Race3)
```



Create a new variable *Race3.reord* to re-order *Race3*
with the categories in frequency order

```
levels(nhanes_1000$Race3)
```

```
nhanes_1000$Race3.reord =  
  factor(nhanes_1000$Race3, levels = c("White",  
    "Black", "Mexican", "Hispanic", "Asian", "Other") )
```

```
iNZightPlot(Race3.reord, data=nhanes_1000)
```

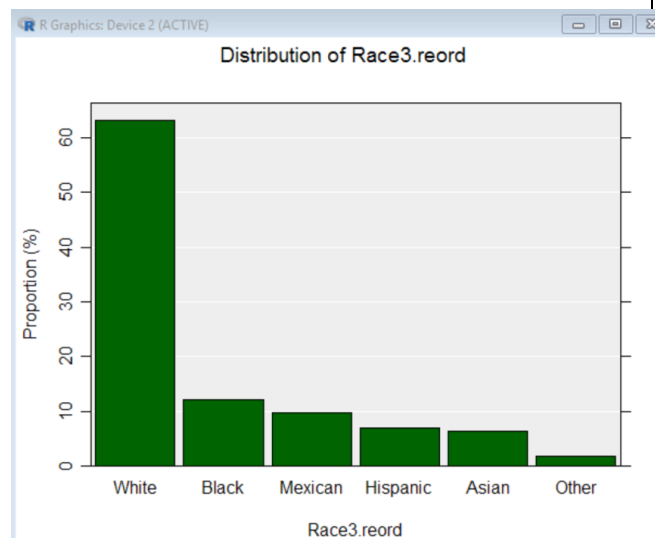
COMMENTARY

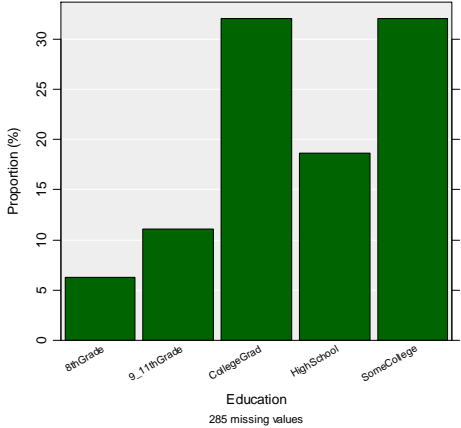
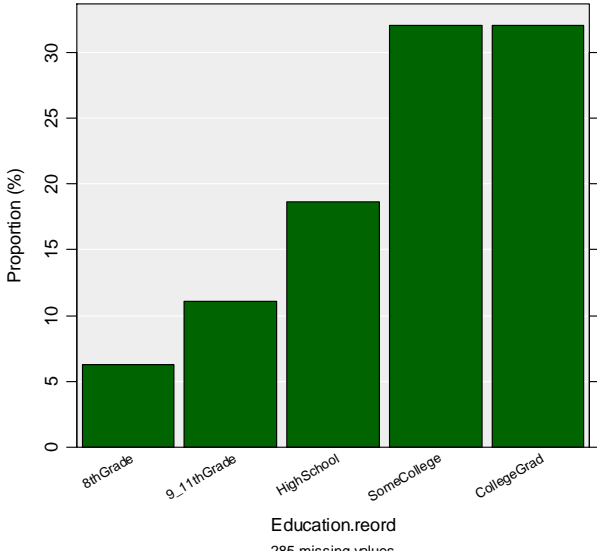
R calls a *categorical* variable a “*factor*”

Show me the *levels* of *Race3* (I can also see in the graph). Output is ...

```
[1] "Asian" "Black" "Hispanic" "Mexican" "Other"  
"White"
```

I can see what the frequency order should be from the graph. (This can be done generally with code but the code is too complex to do at this stage) So I'll make *Race3.reord* from *Race3* and put them in the order I want. (Getting the number of levels and spelling exactly right is crucial)



<p><i># We'll do this again putting the levels of Education into a sensible order</i></p> <pre>iNZightPlot(Education, data=nhanes_1000)</pre> <pre>levels(nhanes_1000\$Education)</pre>	<p>Distribution of Education</p>  <pre>[1] "8thGrade" "9_11thGrade" "CollegeGrad" "HighSchool" "SomeCollege"</pre>
<p><i># Create a new variable to re-order Education</i></p> <pre>nhanes_1000\$Education.reord = factor(nhanes_1000\$Education, levels = c("8thGrade", "9_11thGrade", "HighSchool", "SomeCollege", "CollegeGrad"))</pre> <pre>levels(nhanes_1000\$Education.reord)</pre>	<pre>[1] "8thGrade" "9_11thGrade" "HighSchool" "SomeCollege" "CollegeGrad"</pre>
<pre>iNZightPlot(Education.reord, data=nhanes_1000)</pre>	<p>Distribution of Education.reord</p> 
<pre>iNZightPlot(Education.reord, data=nhanes_1000,colby=Education.reord)</pre> <p><i># Now change the colour palette to rainbow colours</i></p> <pre>iNZightPlot(Education.reord, data=nhanes_1000,colby=Education.reord, col.fun=rainbow)</pre>	<p>COMMENTARY</p> <p><i>Colour by Education.reord</i></p> <p><i>Col.fun has to be a colour palette function</i> <i>There are lots of colour palette functions in R,</i> <i>many you have to install other packages to get.</i> <i>rainbow() is a generally available colour palette</i></p>

<pre>library(colorspace) iNZightPlot(Education.reord, data=nhanes_1000,colby=Education.reord, col.fun=rainbow_hcl)</pre>	<i>Using the rainbow_hcl colour function from the colorspace package</i>
--	--

- Try repeating the above using other choices for variables and settings

If you want to try installing some other R packages, in the R menus

Go **Packages > Install packages** . You will probably be asked to choose a CRAN mirror site.

Then you will be shown a list of packages to choose from.

Installing the package **viridis** and then loading it [via *library(viridis)*] will give you access to the colour functions: **viridis**, **magma**, and **inferno**

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!