



## 4. Categorical variables

### Key iNZight skills addressed:

- Visual displays (Bar and Pareto charts) and summary tables for each categorical variable.
- Colouring the bars.
- Re-ordering the responses (and hence bars) based on their frequency.

### Relevant StatsTuneUp video to consider:

- 'Random Variables' available at <http://statstuneup.com.au/random-variables/>
- 'Visual Displays: Single Variable' available at <http://statstuneup.com.au/one-variable/>

#### Side note: Types of categorical variables

There are two types of categorical variables: **ordinal** and **nominal** variables.

The variable type determines how you display the data.

**Recommended viewing:** please consider the 4-minute animated videos 'Random Variables' and 'Visual Displays: Single Variable' available at <http://statstuneup.com.au/random-variables/> and <http://statstuneup.com.au/one-variable/> (click on 'The Video' tab to view)

#### Ordinal Variables

When there is a natural ordering or hierarchy of the possible responses or *categories* (e.g., the variable 'Risk' having possible responses of Low, Low-Medium, Medium, Medium-High, High) we have an ordinal variable. It is important to keep this order when you present your data graphically, or in a table showing the frequencies of each categorical response. Such a graph is referred to as a bar chart.

#### Nominal Variables

When there is no natural ordering or hierarchy to the possible responses or *categories* (e.g., the variable 'eye colour' having possible responses of brown, grey, green, blue, etc) we have a nominal variable. When we present such a variable in a graph, or in a table showing the frequencies of each categorical responses, there is no particular order for the categories we must retain. The categorical responses may be arranged in descending order of frequency for example. Such a graph is referred to as a Pareto chart.

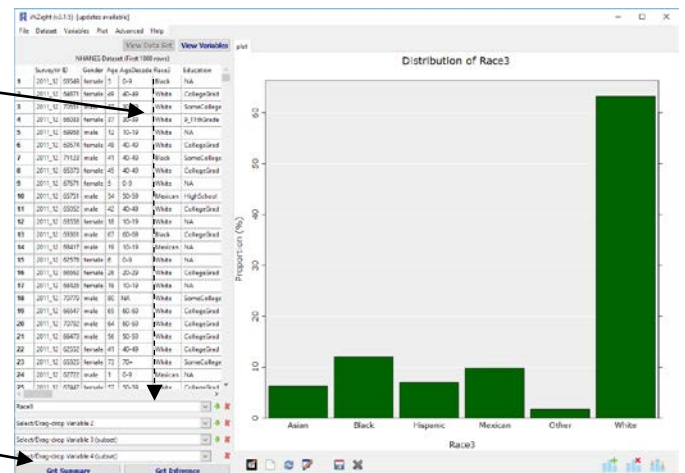
## INSTRUCTIONS

Load the **nhanes\_1000** dataset into iNZight using **File > Example data ...**. You will find the data set in **Module (package) FutureLearn**. See earlier worksheets if unfamiliar with this step.

## Creating bar charts of categorical variables

- Drag the variable name **Race3** into the **Variable1** slot.

- Click **Get Summary** to obtain the associated summary table of counts and percentages.



See pages 5-6 of this document to learn how to copy this output into Excel

iNZight Summary							
Primary variable of interest: Race3 (factor)							
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%

### ASIDE: PRACTICE (~2 min)

Drag the names of some other **categorical variables** into the **Variable 1** slot - ensure you can get simple graphs and tables like the above quickly for each variable.

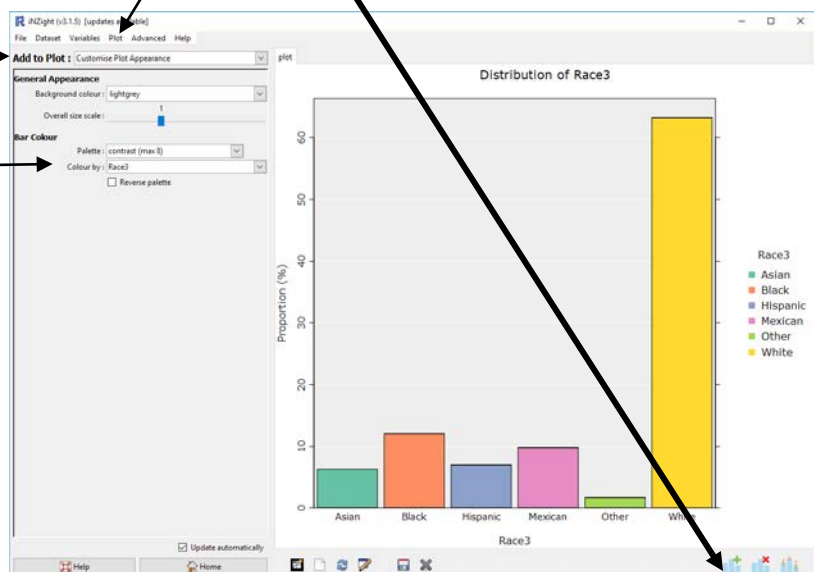
## Colouring the bars

- Having 'Race3' variable in **Variable 1** slot, click on the **Add to Plot** icon located beneath the plot window. (*Add to Plot* can also be found in top menu: *Plot > Add to plot ...*)

- **Customise plot appearance** appears in the first field by default.

- Under **Colour by** choose **Race3**.

Try changing the colour **Palette** to rainbow(hcl).

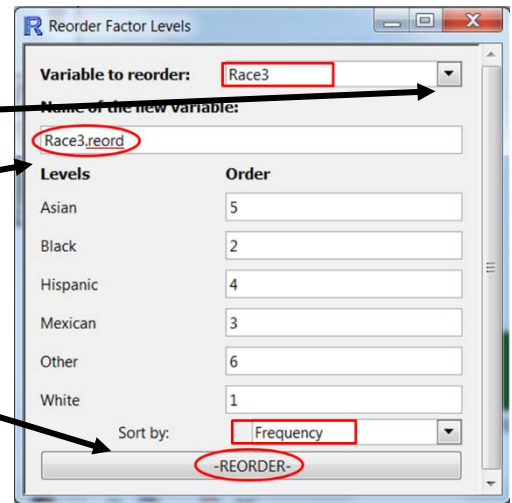


## Reordering the bars by frequency

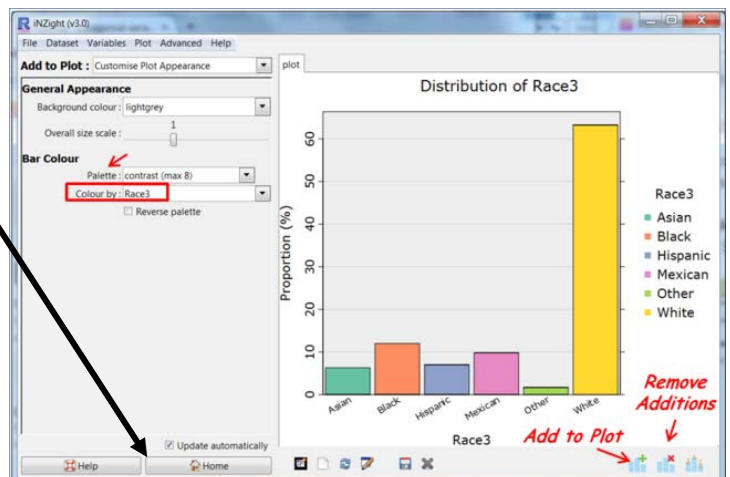
iNZight initially displays the bars of a graph in alphanumeric order. **From the top menu:**

- select **Variables > Categorical Variables > Reorder Levels**
- A new window which appears, select **Race3** from the drop down list beside **Variable to reorder**
- give the variable a new name or accept the name offered, **Race3.reord**
- Select Sort by: **Frequency** and click **REORDER**.

**This does not change the graph.** It just creates a new variable that has the desired ordering – you can then plot this new variable.



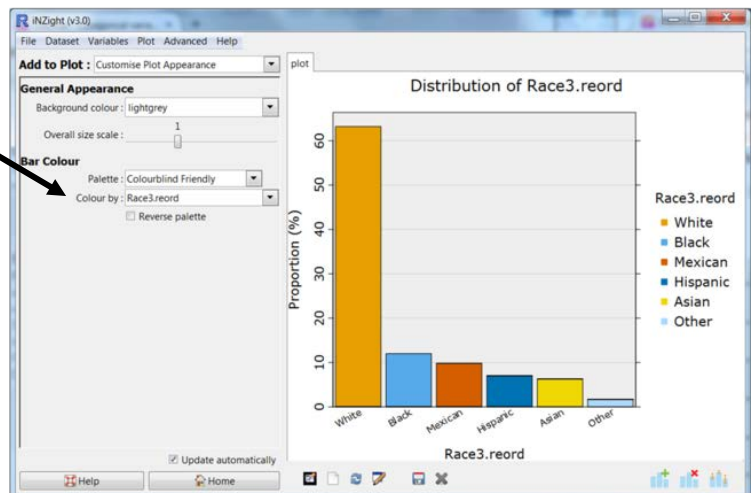
- Select the **Home** button (at bottom) to return to the Home command panel so that you can plot other variables.



- Drag **Race3.reord** (or your new variable name) into the **Variable1** slot.
- Choose colour by **Race3.reord**.

*(Here the so-called Colourblind friendly palette has been nominated.)*

This is an example of a Pareto chart – appropriate for nominal variables.



## Troubleshooting

- ***I resized the window and my plot disappeared.***
  - Click the **Redraw plot** icon (*you can hover over the icons to see what they do*).
- ***How do I remove the colour segmentation in the bars?***
  - You have a colour-by setting that is still active. Go to **Remove Additions** underneath the plot and then use **Remove colour by**.
- ***The plot didn't change when I used reorder categories and clicked "Reorder".***
  - Clicking re-order does not create a new graph, it creates a new variable (e.g. race3.reord).

To get the new (reordered) graph you have to drag/select the new variable.  
[The behaviour of iNZight Lite is different ...](#).
- ***My bar labels are on an angle. How do I make them horizontal?***
  - Stretch the plot window horizontally until there is room for them to become horizontal and then click the **Redraw plot** icon (underneath the plot window) to redraw the plot.

Another option is **Add to plot** and under *Overall Plot Appearance*, reduce **Overall size scale** (makes every plot element smaller).
- ***How do I modify headings?***
  - Use **Add to Plot > Axes and labels**.
- ***How do I save my graph from iNZight for use in another document?***
  - Click on the **Save Plot** icon at the bottom of your graph, or go **Plot > Save Plot**. There several formats available, e.g. pdf, png, ...
- ***How do I copy and paste my graph from iNZight into another document?***
  - Copy and paste only works in Windows. (Otherwise you have to use screen shots.)

Click on the **New graphics window** icon at the bottom of your graph (*you can hover over the icons to see what they do*), then you will have another copy of your graph in a separate graphics window.

Right click on this new window, and **Copy as metafile** or **Copy as bitmap**.
- ***How can I save the data file so that it remembers the reordering I have done?***
  - Selecting **File > Save ...** enables you to save the data in a R native format which remembers all new variables you have created including features like reorderings and renamings - R is the free statistical computing software used to enable iNZight.

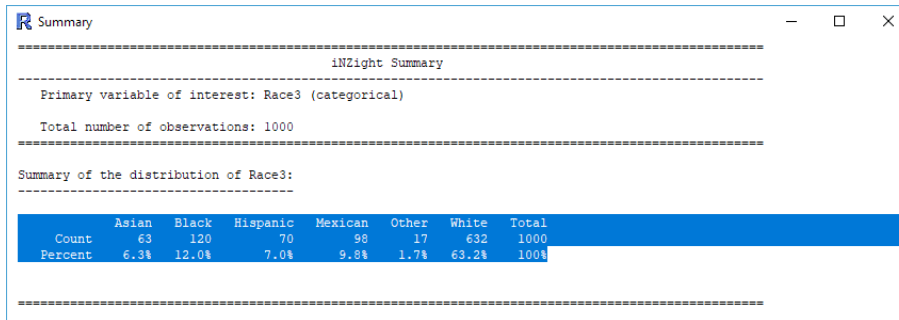
You may then read it back in a new iNZight session using **File > Load ...**.

### Additional:

- for **how to** instructions for iNZight you should consult iNZight's [User Guides](#).
- for **technical issues** with running iNZight, you should first consult iNZight's [frequently asked questions \(FAQ\)](#) page and if the answer you need is not there ask at [iNZight support](#).

## How to copy summary information to Excel

Highlight the summary and copy:



The screenshot shows a window titled "Summary" with the following content:

iNZight Summary

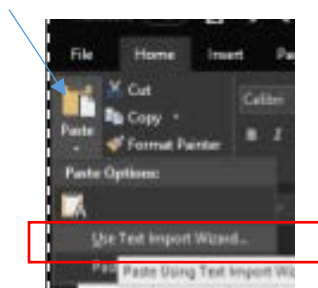
Primary variable of interest: Race3 (categorical)

Total number of observations: 1000

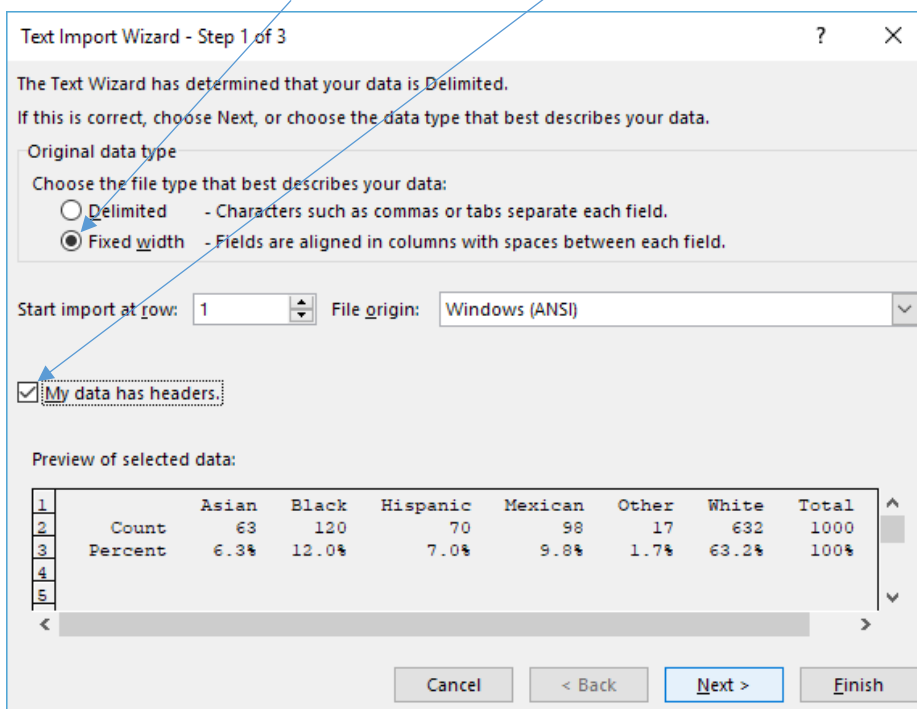
Summary of the distribution of Race3:

	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%

In Excel, click on Paste and select Use Text Import Wizard:



Choose "Fixed width" and "My data has headers"



The screenshot shows the "Text Import Wizard - Step 1 of 3" dialog box. The "Original data type" section has "Fixed width" selected. The "Start import at row:" is set to 1 and "File origin:" is set to Windows (ANSI). The "My data has headers" checkbox is checked. A preview of the data is shown at the bottom.

Text Import Wizard - Step 1 of 3

The Text Wizard has determined that your data is Delimited.  
If this is correct, choose Next, or choose the data type that best describes your data.

Original data type

Choose the file type that best describes your data:

Delimited - Characters such as commas or tabs separate each field.

Fixed width - Fields are aligned in columns with spaces between each field.

Start import at row: 1 File origin: Windows (ANSI)

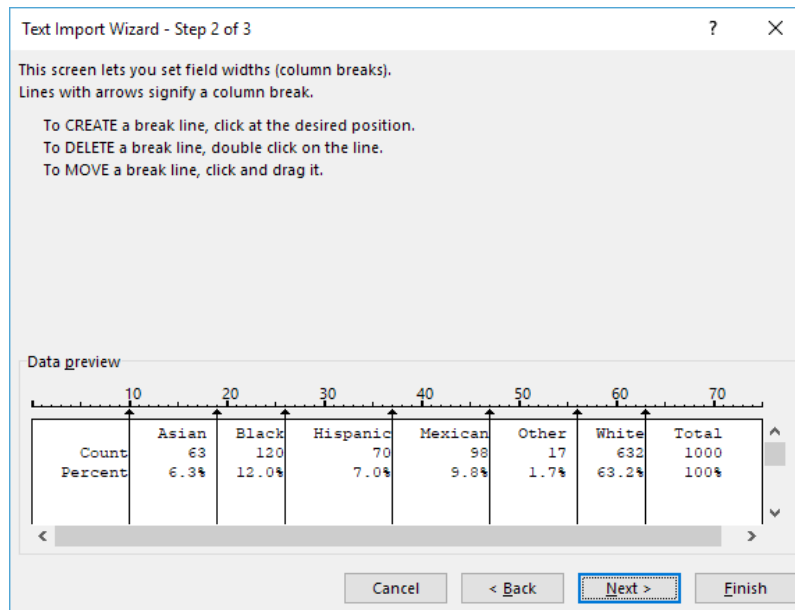
My data has headers

Preview of selected data:

1		Asian	Black	Hispanic	Mexican	Other	White	Total
2	Count	63	120	70	98	17	632	1000
3	Percent	6.3%	12.0%	7.0%	9.8%	1.7%	63.2%	100%
4								
5								

Buttons: Cancel, < Back, Next >, Finish

Click 'Next' – the following should appear



Click 'Next' and 'Finish' – the following should appear

	A	B	C	D	E	F	G	H
1		Asian	Black	Hispanic	Mexican	Other	White	Total
2	Count	63	120	70	98	17	632	1000
3	Percent	6.30%	12.00%	7.00%	9.80%	1.70%	63.20%	100%
4								