6.12 Exercise: Inference with iNZight – *R version*

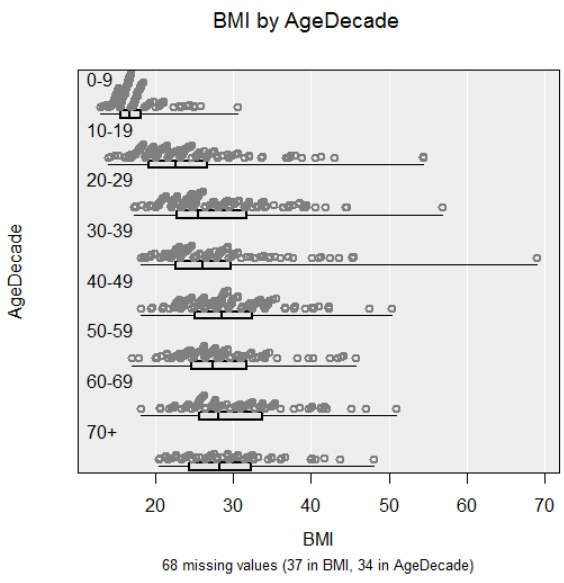
This exercise will enable you to make comparisons between sub-groups allowing for sampling error.

Background understanding: see ***Step 6.9.***

The skills addressed in this Exercise are:

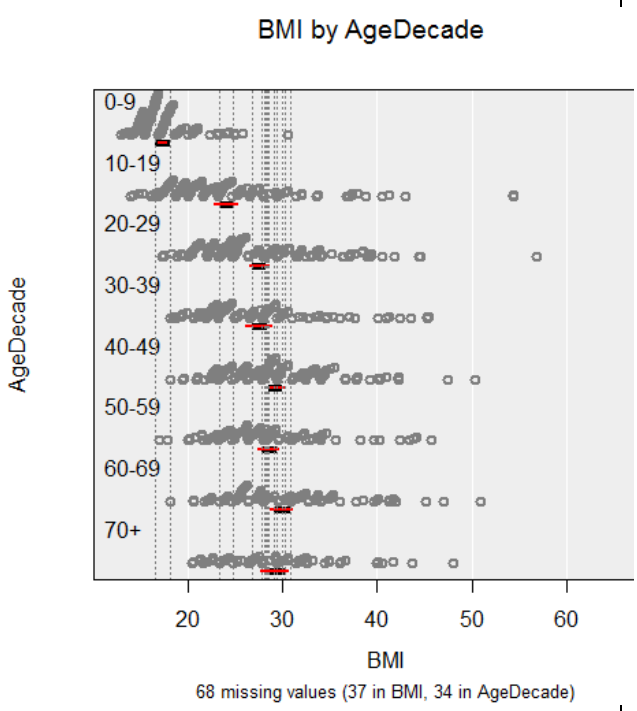
- Use iNZightPlot to get inferential mark-ups of plots so that you can make visual comparisons between sub-groups allowing for sampling error.
- To obtain numerical confidence limits for true between-group differences.

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

#R code	Output and/or Commentary
# Setup library(iNZightPlots) library(FutureLearnData) **data(nhanes_1000)** **names(nhanes_1000)**	
# Plot BMI by AgeDecade iNZightPlot(**AgeDecade, BMI,** data=nhanes_1000**)**	

 # Add inference information

```
iNZightPlot(AgeDecade, BMI, data=nhanes_1000,
        inference.type=c("comp","conf"),
        inference.par="mean")
```



*# **Commentary** (on this side to save space)*

*# You can **squash** your plot window **vertically** so that it is **easier to see** how much **overlap** there is between each age group.*

*# What do you see here? The **thick black lines** are called **'comparison intervals'** and are the lines that we look at when are the lines that we **look at** when observing any **overlap.** The **thin red lines** are the **individual confidence intervals** for each mean/median.*

*# **View detailed inferential information** (Normal Theory)*

```
getPlotSummary(AgeDecade, BMI, data=nhanes_1000,
        summary.type="inference", inference.type="conf")
```

```
================================================================================
                          iNZight Inference using Normal Theory
--------------------------------------------------------------------------------
       Primary variable of interest: AgeDecade (categorical)
             Secondary variable: BMI (numeric)

      Total number of observations: 1000
  Number omitted due to missingness: 68 (34 in AgeDecade, 37 in BMI)
  Total number of observations used: 932
================================================================================

Inference of BMI by AgeDecade:
------------------------------

Group Means with 95% Confidence Intervals

           Lower    Mean   Upper
     0-9   16.86   17.36   17.85
   10-19   22.78   24.05   25.31
   20-29   26.49   27.57   28.64
   30-39   26.16   27.58   29.01
   40-49   28.35   29.28   30.21
   50-59   27.46   28.58   29.71
   60-69   28.74   29.98   31.21
```

etc

We'll now do this with a pair of Categorical variables

*# **Filter out under 20s***

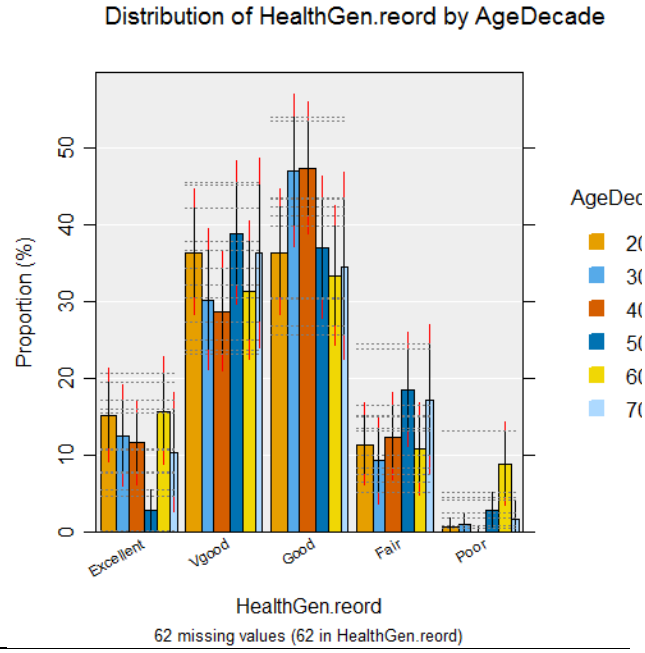
```
Temp=subset(nhanes_1000,AgeDecade !=" 0-9" &
        AgeDecade !=" 10-19")
Temp$AgeDecade=factor(Temp$AgeDecade)
```

*# **Reorder the HealthGen variable***

```
Temp$HealthGen.reord = factor(Temp$HealthGen, levels
        = c("Excellent","Vgood","Good","Fair","Poor") )
```

***First need to do some** filtering to get rid of decades with no General Health data*

*and **create HealthGen.reord** with the levels in a sensible order*

# Plot as separate bar charts with inference information iNZightPlot(HealthGen.reord, **g1=**AgeDecade, data=Temp, **inference.type=c("comp","conf"))**	
# Plot as side by side without inference comparison iNZightPlot(HealthGen.reord, AgeDecade, data=Temp, **inference.type=c("comp","conf"))**	

- **Play some more with these settings and try other variables**

<u>To discuss issues related to this Exercise</u>,

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!