

## 4.11 Exercise: Advanced scatterplots for deeper analysis – R version

Note: *Copying and pasting text (e.g. R code) from a pdf is not reliable. For that reason we have also provided the code in [a text file](#)*

This exercise will enable you to explore more complicated relationships between variables and the explore the effects of a third and fourth variable, enabling you to view changes over time.

The skills addressed are:

1. Create a scatterplot of two numeric variables, subset by a 3rd variable.
2. Explore the effect of a third and fourth variable using colour and size.

We will use the **gapminder** dataset (but **not** *gapminder\_2008*).

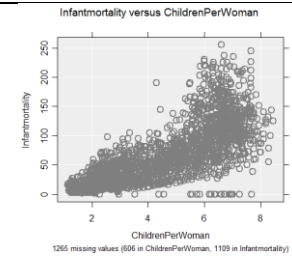
### Create a scatterplot of two numeric variables, subset by a 3rd variable

We are going to explore the relationship between the variables **Infantmortality** and **ChildrenPerWoman** of countries in the **Gapminder** dataset over time.

#R Code	Output and/or Commentary
<pre> # Setup library(iNZightPlots) library(FutureLearnData) data(gapminder) </pre>	

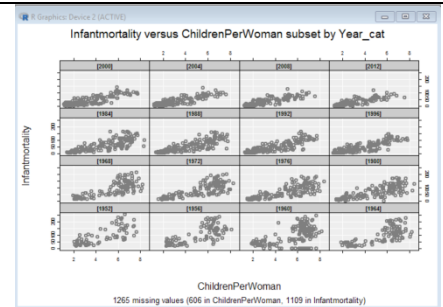
# Scatterplot of *Infantmortality* against *ChildrenPerWoman*

```
iNZightPlot(ChildrenPerWoman,Infantmortality ,  
data=gapminder)
```



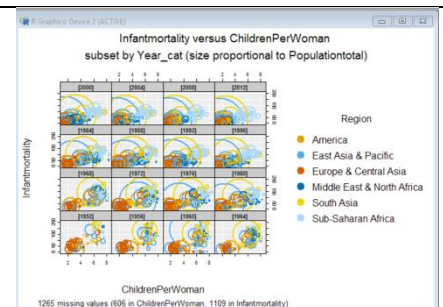
# *Subset by Year\_cat*

```
iNZightPlot(ChildrenPerWoman,Infantmortality, g1=Year_cat,  
data=gapminder)
```



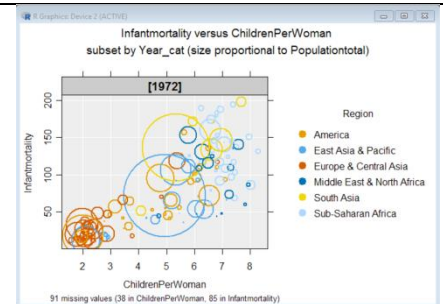
# *Change size and colour of points*

```
iNZightPlot(ChildrenPerWoman,Infantmortality,g1=Year_cat,  
data=gapminder, colby=Region, sizeby=Populationtotal)
```



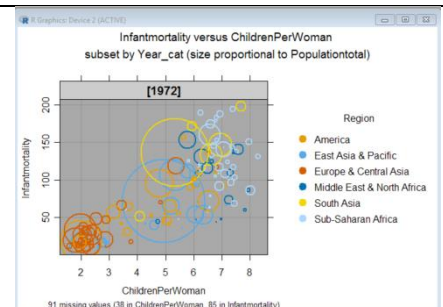
# *Show results for 1972 only*

```
iNZightPlot(ChildrenPerWoman,Infantmortality,g1=Year_cat,  
g1.level="[1972]",data=gapminder, colby=Region,  
sizeby=Populationtotal)
```



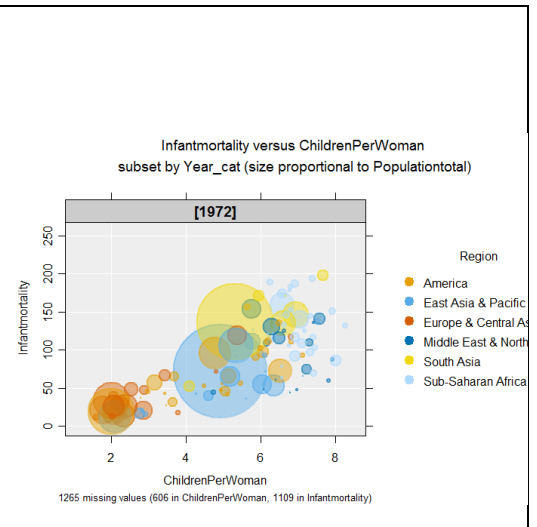
# *Darker background (often easier to see some of the lighter dots)*

```
iNZightPlot(ChildrenPerWoman,Infantmortality, g1=Year_cat,  
g1.level="[1972]",data=gapminder, colby=Region,  
sizeby=Populationtotal, bg="darkgray")
```



# Try transparency and smaller points (removed garkgray)

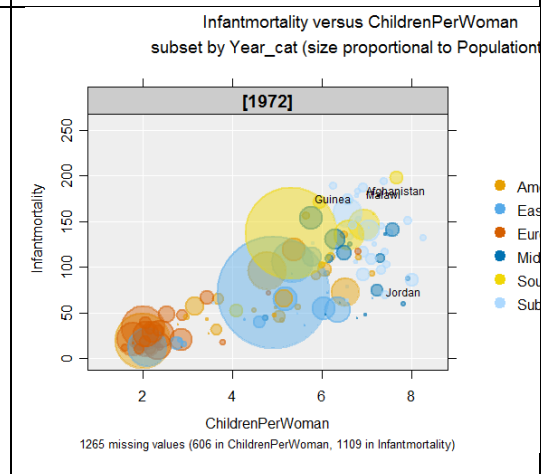
```
iNZightPlot(ChildrenPerWoman,Infantmortality, g1=Year_cat,
  g1.level="[1972]",data=gapminder, colby=Region,
  sizeby=Populationtotal, alpha=.45, cex.dotpt=.5)
```



# Try subsetting by different years, e.g. g1.level="[1976]",

# Label some of the extreme points (ask for 4)

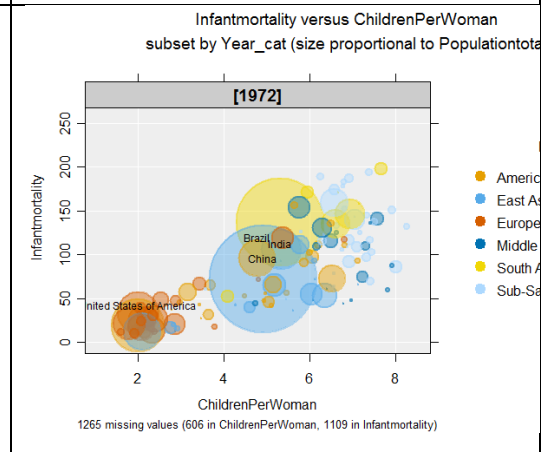
```
iNZightPlot(ChildrenPerWoman,Infantmortality, g1=Year_cat,
  g1.level="[1972]",data=gapminder, colby=Region,
  sizeby=Populationtotal, alpha=.45, cex.dotpt=.5,
  locate.extreme=4, locate=Country)
```



# Label some specific countries

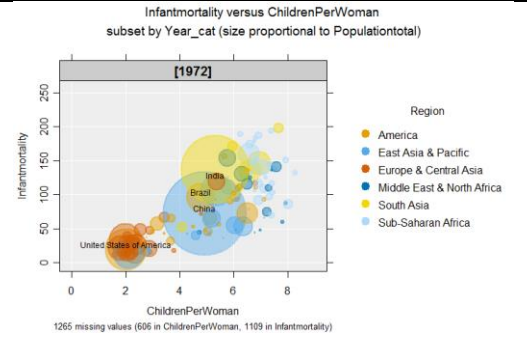
```
ids = (1:nrow(gapminder))[gapminder$Country %in%
  c("United States of America", "China", "Brazil", "India")]
```

```
iNZightPlot(ChildrenPerWoman,Infantmortality, g1=Year_cat,
  g1.level="[1972]",data=gapminder, colby=Region,
  sizeby=Populationtotal, alpha=.45, cex.dotpt=.5,
  locate.id=ids, locate=Country)
```



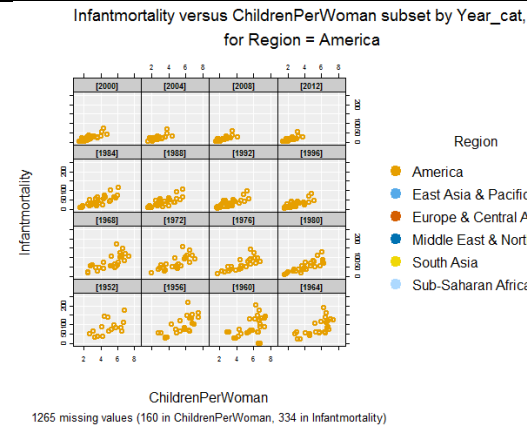
```
# Allow a little more room on left to accommodate label
```

```
iNZightPlot(ChildrenPerWoman, Infantmortality, g1=Year_cat,
             g1.level="[1972]", data=gapminder, colby=Region,
             sizeby=Populationtotal, alpha=.45, cex.dotpt=.5,
             locate.id=ids, locate=Country, xlim=c(0,9))
```



```
# Subset by a fourth variable (Region)
```

```
iNZightPlot(ChildrenPerWoman, Infantmortality, g1=Year_cat,
             g2=Region, g2.level="America", data=gapminder, colby=Region)
```



```
# Play through the years
```

```
for (k in levels(gapminder$Year_cat)) {
  iNZightPlot(ChildrenPerWoman, Infantmortality, g1=Year_cat,
              g1.level=k, data=gapminder, colby=Region,
              sizeby=Populationtotal, alpha=.45, cex.dotpt=.5,
              locate.id=ids, locate=Country)
  Sys.sleep(1)
}
```

- Play some more with these settings and try other variables
- For even more settings, type `?inzpar` into R to get help on the `inzpar`, or type `inzpar` to just get a complete list (last time I looked the help file wasn't entirely complete)

## Optional

If you want to play with a more recent version of the data download [Gapminder 1952-2016 by 4yrs.csv](https://www.stat.auckland.ac.nz/~wild/data/gapminder_datasets/) from [https://www.stat.auckland.ac.nz/~wild/data/gapminder\\_datasets/](https://www.stat.auckland.ac.nz/~wild/data/gapminder_datasets/). [Some variable names may differ somewhat from those in the Exercise above.]

**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!***