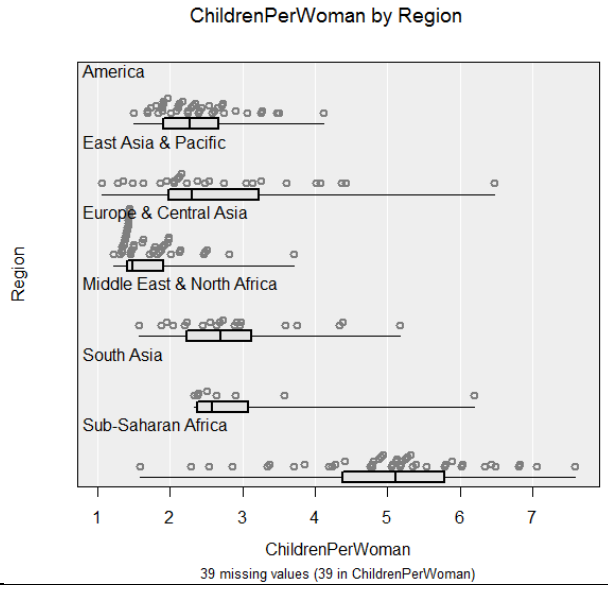
2.13 Exercise: Comparing groups – *R version*

We are going to use the **gapminder_2008** data in the **FutureLearnData** package and look at the values of **ChildrenPerWoman** for each of the countries broken out by **Region**.

# R CODE	COMMENTARY or OUTPUT
# Setup library(iNZightPlots) library(FutureLearnData) **data(gapminder_2008)** names(gapminder_2008) iNZightPlot(**ChildrenPerWoman**, **data=*gapminder_2008*)**	***Commentary*** *Make **gapminder_2008** inside **FutureLearnData** available for analysis* *Useful for checking on the spellings of variable names* *Dot plot for **ChildrenPerWoman***
# Now break out by Region iNZightPlot(ChildrenPerWoman**, Region**, data=gapminder_2008)	

Get Summary of ChildrenPerWoman broken out by Region

getPlotSummary(ChildrenPerWoman**,** Region, data=gapminder_2008**)**

```
-------------------------------------------------------------------------------------------------
        Primary variable of interest: ChildrenPerWoman (numeric)
              Secondary variable: Region (categorical)

      Total number of observations: 225
  Number omitted due to missingness: 39 (39 in ChildrenPerWoman)
  Total number of observations used: 186
=================================================================================================

Summary of ChildrenPerWoman by Region:
-------------------------------------

                              Min    25%  Median    75%    Max   Mean     SD  Sample Size
                America     1.498  1.908   2.264  2.668  4.119  2.377 0.5779           40
        East Asia & Pacific 1.050  1.979   2.295  3.226  6.479  2.685 1.2292           26
     Europe & Central Asia  1.218  1.410   1.482  1.898  3.703  1.713 0.4700           48
Middle East & North Africa  1.570  2.223   2.703  3.125  5.163  2.893 0.9310           20
                South Asia  2.323  2.381   2.574  3.069  6.196  3.113 1.3125            8
        Sub-Saharan Africa  1.579  4.377   5.117  5.785  7.588  4.990 1.2674           44

=================================================================================================
```

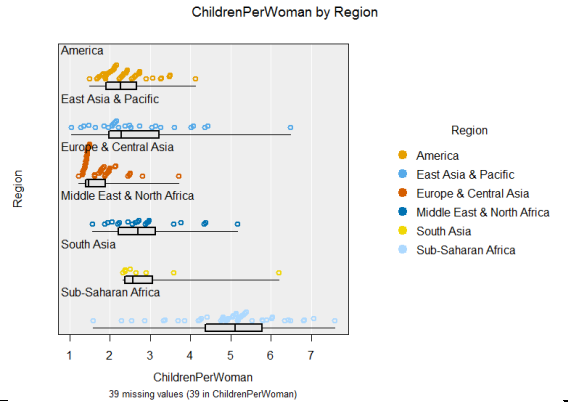
Colour by Region

iNZightPlot(ChildrenPerWoman, Region, data=gapminder_2008, **colby=Region**)

Try also

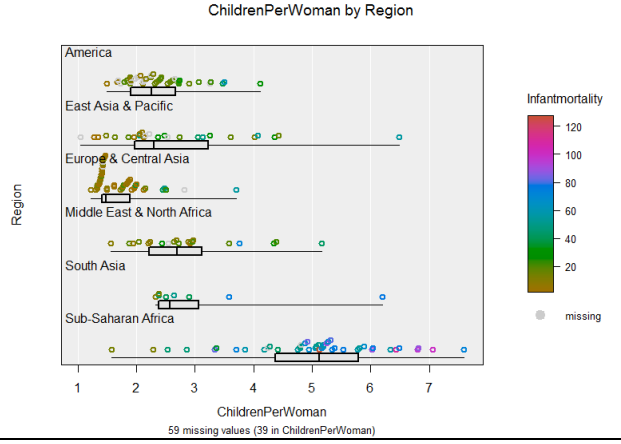
iNZightPlot(ChildrenPerWoman, Region, data=gapminder_2008, colby=Region, **cex.text=.3**)

iNZightPlot(ChildrenPerWoman, Region, data=gapminder_2008, colby=Region, **hide.legend = TRUE**)



Colour by Infantmortality

iNZightPlot(ChildrenPerWoman, Region, data=gapminder_2008, **colby=Infantmortality**)



- **What do you see in the last graph?**

- **Also try colouring by other variables you think might help explain the Regional differences.**

<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!