

2.13 Exercise: Comparing groups (iNZight Lite version)

This exercise will enable you to construct the types of graphs and summaries used in the previous video. The skills addressed are:

- 1. Get dot plots and summary tables when both a numeric variable and a categorical variable (containing the group labels) are used.
- 2. Colour by a third variable.

INSTRUCTIONS

Follow these instructions to generate the plots. If you have a problem doing the exercise, see the common questions on page 5.

Import the Gapminder-2008 dataset into iNZight Lite:

- Select File > Dataset Examples
- Select Data set category: Future-Learn
- Select Gapminder-2008
- Click on Select Set
- Click Visualize

Make dot plots from one numeric and one categorical variable

Select the variable name **ChildrenPerWoman** from the **First variable** slot and **Region** from the **Second variable** slot. iNZight automatically produces a set of dot plots for comparing groups whenever the name of a **numeric** variable is dragged into the **First variable** slot and the name of a **categorical** variable is dragged into the **Second variable** slot.

Select Variables Add To Plot	Pl	ot Summary Inference Interactive Plot (Beta)
ariable selection	Plots	i for visualizing data. ChildrenPerWoman by Region
childrenPerWoman		America o <i>がの見coth Both con 見</i> o o の の の
Region	Region	0 00 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
none v		0 0 0 0 0 0 South Asia 0 0 0 0 0 0 Sub-Saharan Africa 0 0 0 0 0 0
none ×		0 0
A REVERT TO OLD VERSION	±.	DOWNLOAD PLOT Select the file type
RESET TO DEFAULT		

Generate the Summary window

Click on the **Summary** tab to view the associated summary table.

Region Image: Coll number of observations: 2925 Subset by: Summary of ChildrenPerkkoman by Region: none Image: Coll number of observations Min 25% Median 75% Max Menica 1.498 Lings 2.226 America 1.498 Lings 1.998 Lings 6.457 Lings 1.998 Lings 6.457 Lings 1.998 Ling	Select Variables Add To Plot	Plot Summary Inference	e Int	eractive	Plot (Beta)					
Select first variable: INZight Summary ChildrenPerWoman Select second variable: Region Number of tobservations: Subset by: none none v	/ariable selection	Statistical Sumary for the data.									
ChildrenPerWoman Primary variable of interest: ChildrenPerWoman (numeric) Secondary variable: Region (categorical) Select second variable: Total number of observations: 225 Number omitted due to missingness: 39 (39 in ChildrenPerWoman) Total number of observations: used: 186 Subset by: Summary of ChildrenPerWoman by Region: Subset by: Sumary of ChildrenPerWoman by Region: Subset by: Min 25% Median 75% Max Mean 5D Sample Size Region (categorical 1.498 1.988 2.264 2.668 4.119 2.377 0.5779 40 East Asia & Porific: 1.989 1.909 2.925 3.226 6.479 2.265 1.2202 26 Europe & Central Asia 1.218 1.410 1.482 1.988 3.703 1.713 0.4700 48 Middle East & North Africa 1.578 2.223 2.238 3.125 5.163 2.893 0.9310 20 South Asia 1.238 1.410 1.482 1.574 3.069 6.196 3.113 1.31325 8	Select first variable:										
ChildrenPerWoman Selet second variable:						-					
Region Image: State in the original subservations: 29 (39 in ChildrenPerkoman) Subset by: Total number of observations used: 186 subset by: Summary of ChildrenPerkoman by Region: subset by: Subset by: none Image: State in the original subset	ChildrenPerWoman	Primary variable of int	terest: (Children	PerWoman	(numeric					
Region Number omitted due to missingness: 39 (39 in ChildrenPerWoman) Total number of observations used: 186 Subset by: Summary of ChildrenPerWoman by Region: Subset by: Manerica 1.498 1.098 2.264 2.668 4.119 2.377 0.5779 40 Subset by: East Asia & Parific 1.696 1.579 2.229 3.225 6.479 2.654 1.4713 0.4700 48 None Midel East & North Africa 1.579 2.223 2.381 3.225 5.163 2.993 0.9119 20	Select second variable:										
Kegion Total number of observations used: 186 Subset by: Summary of ChildrenPerWoman by Region: none Image: Subset by:					- Children	Destination					
Subset by: Min 25% Median 75% Max Mean 5D Sample Size Subset by:	Region 🗸 🔸				1 Childrei	nPerwoman	1)				
Subset by: Min 25% Median 75% Max Mean 5D Sample Size none Min 25% Median 75% Max Mean 5D Sample Size Subset by: Marci La 1.498 1.089 2.264 2.668 4.119 2.377 0.5779 40 Subset by: East Asia & Pacific 1.696 1.597 2.295 3.226 6.479 2.665 1.2292 26 none Middle fast & North Africa 1.578 2.232 2.783 3.225 5.163 2.493 0.9110 20 none Middle fast & North Africa 1.578 2.233 2.381 2.574 3.696 6.196 3.113 1.3125 8											
none Min 25% Median 75% Max Mean 5D Sample Size Subset by: East Asia & Parific 1.498 1.998 2.264 2.668 4.119 2.377 0.5779 40 Subset by: East Asia & Parific 1.690 1.979 2.265 3.226 6.479 2.665 1.2922 26 none Middle East & North Africa 1.570 2.223 2.781 3.125 5.13 2.693 0.9110 20		Summary of ChildrenPerWoman by	Region:								
America 1.498 1.998 2.264 2.668 4.119 2.377 0.5779 40 Subset by: East Asia & Parifiti 1.050 1.979 2.295 3.226 6.479 2.685 1.2292 26 Subset by: Europe & Central Asia 1.218 1.410 1.482 1.989 3.703 1.713 0.4700 48 none Middle East & North Africa 1.570 2.223 2.781 3.125 5.163 2.693 0.919 20	Subset by:										
America 1.498 1.998 2.264 2.668 4.119 2.377 0.5779 40 Subset by: East Asia & Parifiti 1.050 1.979 2.295 3.226 6.479 2.685 1.2292 26 Subset by: Europe & Central Asia 1.218 1.410 1.482 1.989 3.703 1.713 0.4700 48 none Middle East & North Africa 1.570 2.223 2.781 3.125 5.163 2.693 0.919 20	none 🗸		Min	25%	Median	75%	Max	Mean	SD	Sample Size	
Subset by: Europe & Central Asia 1.218 1.410 1.422 1.898 3.703 1.713 0.4700 48 none Middle East & North Arfsica 1.579 2.223 2.783 3.125 5.931 3.125 8		America									
none Middle East & North Africa 2.323 2.381 2.574 3.069 6.196 3.113 1.3125 8		East Asia & Pacific	1.050	1.979	2.295	3.226	6.479	2.685	1.2292	26	
none ~ South Asia 2.323 2.381 2.574 3.069 6.196 3.113 1.3125 8	Subset by:						3.703			48	
SUUCH ASIA 2.323 2.361 2.374 3.609 0.190 5.115 1.3125 0											
	none										

PRACTICE (~5 min)

From the **Gapminder-2008** dataset, choose other variables (one numeric variable and one categorical) that are of interest to you. When you select a different variable into the variable slot it automatically overwrites the previous variable.

Colour by a third variable

Recreate the first graph of **ChildrenPerWoman** and **Region**.

then

- Using the Add to Plot tab click on Point Colour to reveal options
- Colour by InfantMortality.

<u>。 ﷺ inzight </u>	lite About File - Visualiz	e Dataset + Manipulate variables + Advanced +	
Select Variabl	es Add To Plot	Plot Summary Inference Interactive Plot (Beta)	
Customise Plot	Appearance 👻	Plots for visualizing data.	
General Appeara	sce	ChildrenPerWoman by Region	
Plot type:	default 👻	America o_ <i>ii 00<mark>6_ii tout000 cou</mark>∯</i> oo 9 oo_o	Infantmortality
Background colour:	grey93 👻	East Asia & Pacific 0 00 0 0 00 00 00 00 00 00 00 00 00 00	- 120 - 100
Overall size scale:		Europe & Gentral Asia S Middle East & North Africa	- 80
Doint Size		a. 0 000 @ 0 00 @ 0 0 00 0 South Asia	- 40 - 20
Point Colour		0 <u>8000</u> 0	
	Colour interior	Sub-Saharan Africa	missing
Transparency:	300		1
		ChildrenPerWoman 59 missing values (39 in ChildrenPerWoman)	
Colour by:	Infantmortality	& DOWNLOAD PLOT Select the file type	
Colour palette:	viridis magma	● jpg ○ png ○ pdf ○ svg	
	Reverse palette		
	Use Ranks		

PRACTICE (~5 min)

Try some other choices for the numeric variable in the **First variable** slot and the colour-by variable.

Optional

Try this new feature (interactive graphics)



Click on the **Interactive Plot** tab. This will give you an interactive version of your graph that lets you query it in various ways like hovering over points or clicking them. You have to Click **Produce Plot** for it to appear. This is a protection as the conversion process can be slow if there are many points to be drawn.

You can download these plots as Interactive HTML files which you can give to others. They do not need to be connected to iNZight Lite to work.

Other ways of representing data like this

There are several ways of plotting the data like this. Go to Add to Plot and look at what is delivered by the various options under Plot type. Can you see relationships between the ways the various types of graph represent the information? Play with some of the controls for each plot type.

Common questions

Some columns of dots look like they are suspended in mid-air.

iNZight stacks points that are close together but not identical. This problem occurs when the points are discretely spaced but too big to drop down onto the axis. Either stretch the plot window horizontally until there is room for them to drop down or make the points smaller using option: **Change Plot Appearance** and moving the **Adjust size** slider.