

Customized Graphics with Deducer

Dr. Ann Hess
Graybill Statistical Laboratory
Department of Statistics

<http://www.stat.colostate.edu/~hess/Deducer/>

Outline

- Introduction
- Example 1: Do people lie about their weight?
- Example 2: MPG
- Example 3: Why do some mammals have large brains for their size?
- Example 4: Yarn Breaks
- Example 5: Rose Chart
- Saving Work

2

R and Deducer

- R is a free, powerful, command line programming language for statistical computing.
- Deducer (developed by Ian Fellowes) is an R based program, but does not require command line programming.
- The menu based system allows users to quickly create customized graphics and run standard analyses (t-tests, regression, etc).
- As we run analyses using the pull down menus, R code is generated.
- The Deducer website has installation instructions, video examples and access to a Help/Discussion group: www.deducer.org.
- Another (user friendly option) for using R is "Rcmdr" (developed by John Fox).

3

ggplot2

- An nice feature of Deducer is that it creates graphics using the "ggplot2" package from Hadley Wickham.
- ggplot2 is named after the book "Grammar of Graphics" by Leland Wilkinson.
- ggplot2 creates graphs in layers.
- From the Deducer manual: "The huge advantage of the grammar for a GUI is that though there are a finite number of components, almost any possible statistical graphic can be produced. Leading to a hugely flexible tool. With this flexibility comes the cost of complexity."

4

ggplot2 layers

- Geometric Elements represent what you actually see on the plot: points, lines, etc.
- Statistics allow data summarization. This can be as simple as binning, summing or averaging or a formal model.
- Scales include axes and legends.
- Facets breaks up the data into subsets.
- Coordinates changes the coordinate system (e.g. polar, and map projections)
- Other

5

Getting Started with R and Deducer

1. Download and install R (current version is R 3.1.2) *Only first time!
Go to: www.r-project.org
Click on "download R" then select a CRAN mirror.
Download R (binaries for base distribution) for your operating system.

2. Open R and enter: *Only first time!

```
install.packages(c("JGR", "Deducer", "DeducerExtras"))
```

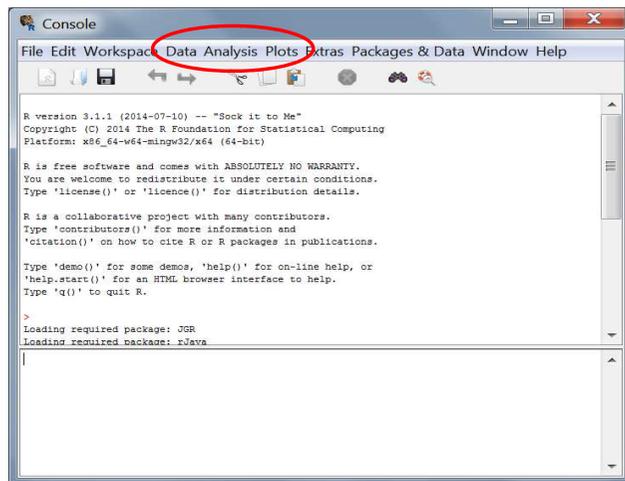
3. In R enter:

```
library(Deducer)  
JGR()
```

The JGR console should now be open.

4. To load Deducer, go to 'Packages & Data' > 'Package Manager' and select Deducer and DeducerExtras. In addition to the standard menus, upon loading Deducer, three new menus should appear named 'Data', 'Analysis' and 'Plots.'

6



7

Importing and Viewing Data

- Standard file types are supported (CSV, TXT, XLSX and others)
- Go to File > Open Data.
- Select the file to be loaded from the file system, and click 'Open'.
- The R name which this data is loaded to can be changed by entering a value into the "Set Name" field at the bottom of the window.
- You may need to change the "Record Separator" depending on your file type. For CSV files, the Record Separator needs to be Comma.
- You should always check the data to make sure it has been read in correctly!
- Go to Packages & Data > Data Viewer. There are tabs for Data View and Variable View (similar to JMP and SPSS).

8

R Packages

- R comes with a standard set of packages, but many more (thousands!) are available for download.
- Before using a library for the first time, you may need to install it. This step only needs to be done once! If a library is already installed, you will still need to load it during the R session when you want to use it. We can load packages using the Deducer Package Manager.
- In addition to the Deducer package, we will also use the “Sleuth2” and “plyr” packages. To install these packages, type (in the lower window):

```
install.packages(c("Sleuth2", "plyr"))
```

9

Example 1: Do people lie about their weight?

- The “Davis” data gives the measured heights (cm) and weights (kg) as well as reported heights and weights for 200 individuals. We also have information about the sex of the subjects (M or F).
- The data is available in R, so to access it we just type (in the lower window):

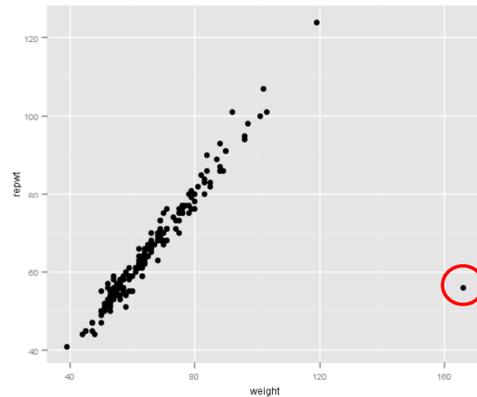
```
data(Davis)
```

- We will cover:
 - Scatter plot, Histogram and Boxplot
 - Summary Statistics
 - Paired t-test and Two-sample t-test

10

Example 1: Scatterplot (Quick)

- Go to Plots > Quick > scatter.
- Choose X = weight, Y = repwt and hit Run.
- We see a strong linear relationship, but there is also an obvious outlier.



11

Example 1: Sorting and Editing Data

- To find the outlier, it is helpful to sort the data by weight.
- Go to Data > Sort. Sort data by: weight and hit OK.
- The outlier is observation #12. It looks like a data entry error, so we will just edit the data before further analysis. Reverse the values for weight and height, so that weight=57 and height=166.
- Another option would have been to remove that observation completely. This can be done by highlighting the row, right-clicking and selecting Remove Row.
- NOTE: In practice, thought should be given to editing or removing data!

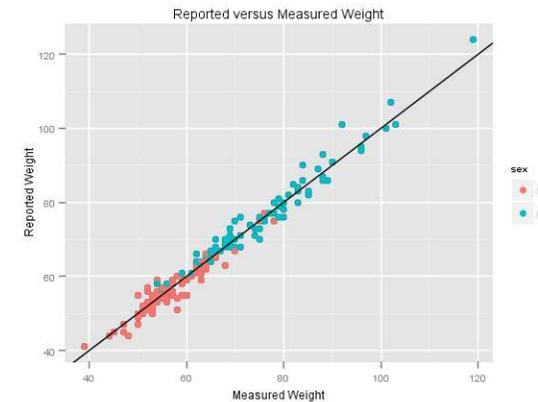
12

Example 1: Scatterplot (Custom)

- Go to Plots > Plot Builder.
- **Scatterplot color coded by Sex:** From Geometric Elements, choose point. Choose X = weight, Y = repwt. Toggle Colour and select sex.
- NOTE: We can control the transparency of the points with Alpha.
- **Add a reference line:** From Geometric Elements, choose abline and specify intercept = 0 and slope = 1.
- **Add a title:** From Other, choose title.
- **Change scale range and labels:** From Scales, choose x (or y) continuous. We can change the labels and also the range.
- When you are happy with the graph hit Run.

13

Example 1: Scatterplot (Custom)



14

Example 1: Creating New Variables

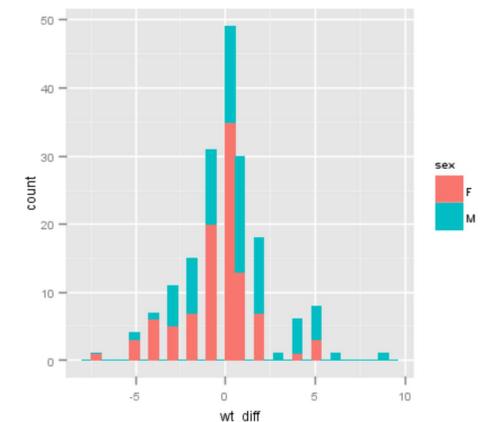
- We are interested in the difference between reported weight versus measured weight.
- We will create a new variable that gives the difference. In practice, this could also be done using Excel.
- To create a new variable in Deducer, we type (in the lower window):

```
Davis$wt_diff = Davis$repwt - Davis$weight
```
- Since we have modified the data, it is always a good idea to check the revised data set.

15

Example 1: Histogram (Custom)

- Go to Plots > Plot Builder.
- **To start a new graph, hit Reset.**
- **To create a basic histogram:** From Templates choose histogram. Choose variable = wt_diff and hit OK.
- **Identify M and F:** From Statistics, choose bin. X = wt_diff and Y = ..count.. Toggle Fill and select sex.
- When you are happy with the graph, hit Run.



16

Example 1: Summary Statistics and t-tests

- Go to Analysis > Descriptives. Choose Descriptives of wt_diff and hit Continue. The mean and median are both close to zero.
- We can run a paired t-test comparing reported versus measured weight. Go to Analysis > Paired test. Choose First = repwt, Second = weight and hit Run. The p-value is high (0.9746) indicating that there is not a significant difference.
- NOTE1: To get the t-test you need to check the box. Wilcoxon signed rank test is a non-parametric alternative to the paired t-test.
- NOTE2: The paired t-test is equivalent to running a one-sample t-test on wt_diff.

17

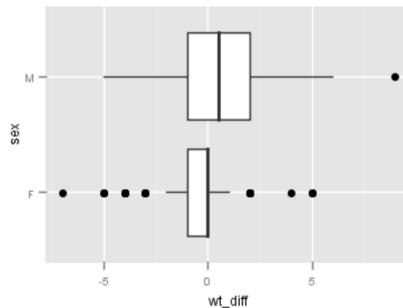
Example 1: Summary Statistics and t-tests

- We can also rerun the summary statistics this time stratifying by sex.
- We see that women tend to under report (negative difference) while men tend to over report (positive difference)!
- We can compare the mean difference for males and females using a two-sample t-test. Go to Analysis > Two Sample Test. Choose Outcome = wt_diff, Factor = sex. The p-value (0.002) is less than alpha=0.05, indicating a significant difference between males and females!

18

Example 1: Boxplot (Quick)

- Go to Plots > Quick > grouped boxplot.
- Choose Variable = wt_diff, Group = sex and hit Run.
- To change the orientation (vertical instead of horizontal), we would need to use the Plot Builder option.



19

Example 2: MPG

- The “mpg” data contains fuel economy data from 1999 and 2008 for 38 popular models of car. We will focus on the variables displ (engine displacement in liters), cyl (# cylinders) and hwy (highway mpg).
- The data is available in R, so to access it we just type (in the lower window):

```
data(mpg)
```

- We will cover:
 - Scatter plot with regression line overlaid
 - Faceting
 - Linear Regression and Subsetting

20

Example 2: Reformatting a variable

- Go to Packages & Data, Data Viewer and go to the Variable View tab.
- Notice that cyl is defined as type “Integer”. However, we would like to treat it as a factor (for our graphs and analysis).
- Simply change the type to “Factor”.

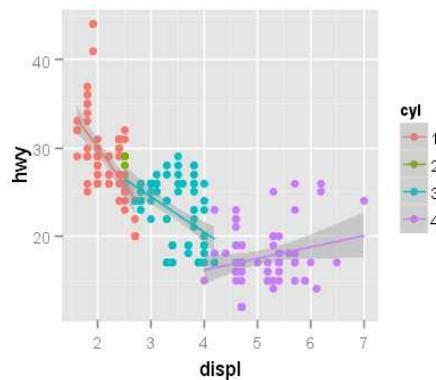
21

Example 2: Scatterplot with Regression lines

- Go to Plots > Plot Builder.
- **The dataset can be changed in the upper left hand corner of a Geometric Elements dialogue window.**
- To create the scatterplot: From Geometric Elements, choose point. X = displ, Y = hwy. Toggle Colour and choose cyl.
- **To add the regression lines (separately for each cyl group):** From Statistics, choose smooth. X = displ, Y = hwy, Colour = cyl, Group = cyl. Under Options, method = lm : Linear.

22

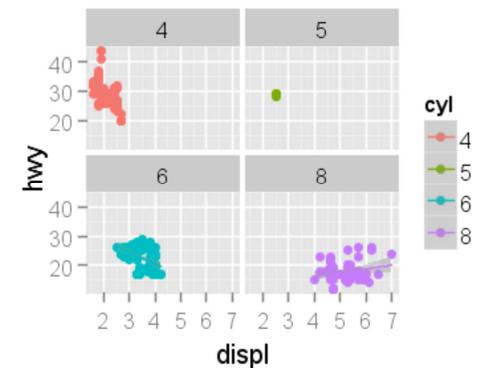
Example 2: Scatterplot with Regression lines



23

Example 2: Faceted scatterplots

- **To create separate graphs for each cyl group:** From Facets choose wrap and facet by = cyl.



24

Example 2: Linear Regression and Subsetting

- Go to Analysis > Linear Model. Outcome = hwy, As Numeric = displ.
- To Run the analysis for only 4 cylinder vehicles, in the Subset box type `cyl==`4``. We see that 4 cylinder vehicles have a negative relationship between displ and hwy (slope = -8.295).
- When we repeat the same (subsetting) analysis for 8 cylinder vehicles, we see that they have a positive relationship between displ and hwy (slope = 1.2964).
- NOTE1: When subsetting, it is a good idea to check the df to make sure the subsetting is working the way you think.
- NOTE2: Click on the Diagnostics tab to see the diagnostic plots for the model. Try the “Pop Out” button!

25

Example 3: Why do some mammals have large brains for their size?

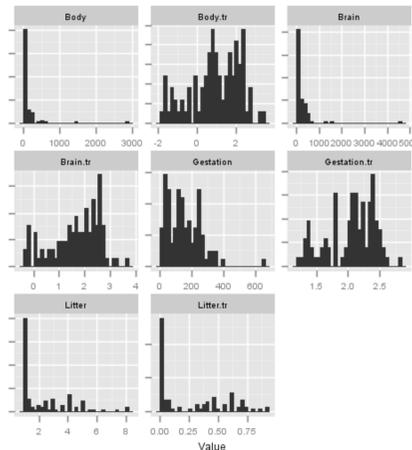
- The “case0902” data gives average values of brain weight, body weight, gestation lengths (length of pregnancy) and litter size for 96 species of mammals. Taken from the Statistical Sleuth by Ramsay and Schafer.
- The data is available from the Slueth2 library. Go to Packages & Data, Package Manager and load “Slueth2”. Then type (in the lower window):

```
data(case0902)
```
- We will cover:
 - Transforming and Renaming Variables
 - Customizing Axes and Adding Labels
 - Bubble Plot
 - Correlation and Multiple Regression

26

Example 3: Transforming Variables

- Following the analysis in Statistical Sleuth, we will log transform all of the variables.
- Go to Data > Transform. Under Variables to Transform select, Brain, Body, Gestation and Litter. Under Transformation, choose Custom $\log_{10}(x)$. Select Plot then Histogram. Hit Run.



27

Example 3: Renaming Variables

- By default the transformed variables are named Body.tr, Brain.tr, Gestation.tr, and Litter.tr.
- To rename the variables, go to Packages & Data > Data Viewer. The names can be changed from Variable View.
- I will use the names Body.log10, Brain.log10, Gestation.log10 and Litter.log10.

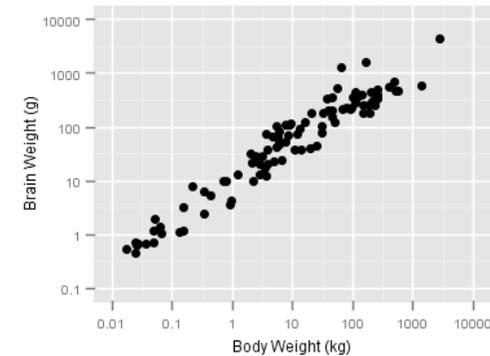
28

Example 3: Scatterplot with log10 axes

- Go to Plots > Plot Builder. From Geometric Elements choose Point. X = Body, Y = Brain.
- Approach #1: From Coordinates choose Trans. X-axis transformation = log10, Y-axis transformation = log10.
- **Approach #2:** From Scales, choose x continuous. Enter values (and labels) of 0.01, 0.1, 1, 10, 100, 1000, 10000. Under Transformation choose log10. Repeat this process for y continuous.
- Approach #3: Plot Brain.log10 vs Body.log10 and just change the axes.

29

Example 3: Scatterplot with log10 axes



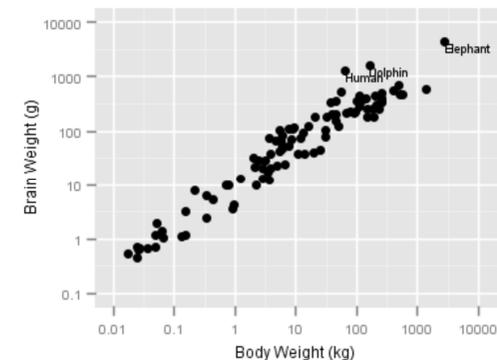
30

Example 3: Adding Text Labels

- In this case, we might want to identify a few species with the largest brain sizes.
- To do this, we will first sort the data by brain size.
- Then create a new variable called "Text". Go to Packages and Data > Data Viewer. Using Variable View create a new Variable called Text of Type Character. When we return to Data View, we should have a column of "NA" values. Type in the labels Human, Dolphin and Elephant for those species only.
- **To add text labels to the graph:** From Plot Builder add Geometric Element Text with Label = Text. The size and position of the text can be adjusted using Size, Hjust and Vjust.

31

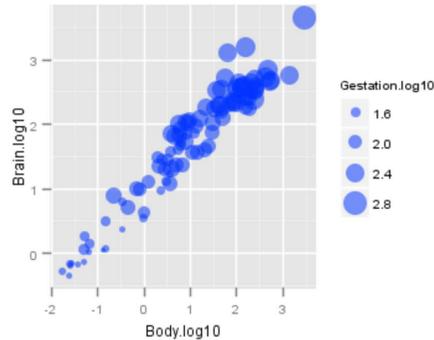
Example 3: Adding Text Labels



32

Example 3: Bubble Plot

- Go to Plots > Plot Builder.
- From Geometric Elements choose Point. X = Body.log10, Y = Brain.log10. Toggle Size and select Gestation.log10.
- **To change the color:** click on Colour: Set colour.
- **Recall that you can adjust the transparency with Alpha.** For this graph, I used Alpha =0.5.



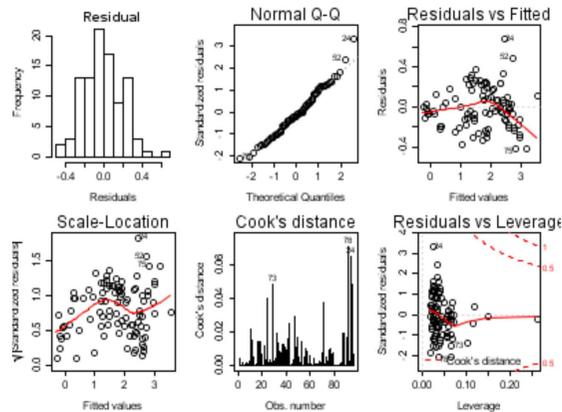
33

Example 3: Correlation and Multiple Regression

- Correlation: Go to Analysis > Correlation. Select all four log transformed variables under Variables.
- We see that Body.log10 and Gestation.log10 are positively correlated with Brain.log10. Litter.log10 is negatively correlated with Brain.log10.
- Multiple Regression: Go to Analysis > Linear Model. Outcome = Brain.log10, As Numeric: Body.log10, Gestation.log10, Litter.log10.
- We see that all three variables are significant predictors of Brain.log10. Model $R^2 = 0.9537$.

34

Example 3: Regression Diagnostic Plots



35

Example 4: Yarn Breaks

- The “warbreaks” data gives the number of warp breaks per loom, where a loom corresponds to a fixed length of yarn. There are two factors: wool (A or B) and tension (H, M or L).
- The data is available in R, so to access it we just type (in the lower window):

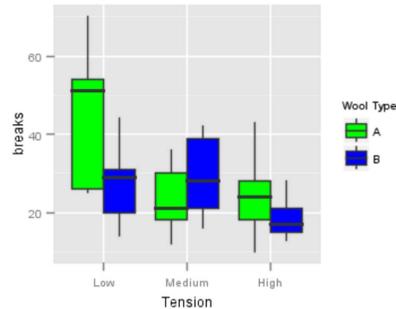
```
data(warbreaks)
```

- We will cover:
 - Grouped Boxplot, Bar Chart and Line Plot
 - Changing Theme, Color and Font Size
 - Two way ANOVA

36

Example 4: Boxplots (two factors)

- Under Geometric Elements, choose boxplot. Y = breaks, X = tension. Toggle Fill and choose wool.
- **To change the colors:** From Scales, choose fill manual. Under values, type green and blue. Can also change the legend here.
- **To edit the labels:** From Scales choose X discrete. Values: L, M, H. Labels: Low, Medium, High.
- **NOTE:** The Included Levels box be used to change the ORDER of the categories!



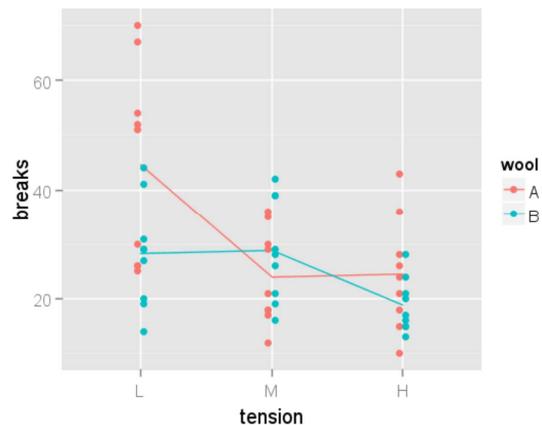
37

Example 4: Line Plot (two factors)

- Go to Plots > Plot Builder.
- **To create a line plot:** Under Geometric Elements, choose line. X = tension, Y = breaks, Colour = wool, Group = wool. Instead of identity, choose summary. Look at options, but the default should be OK.
- **To add a scatter plot:** Under Geometric Elements, choose point. X = tension, Y = breaks, Colour = wool, Group = wool. Position (top left), choose dodge with width = 0.1.

38

Example 4: Line Plot (two factors)



39

Example 4: Summary Statistics

- We can get some basic summary statistics using Analysis > Descriptives and choosing Descriptives of: breaks, Stratify By: wool, tension.
- To get summary plots, we need the means and standard errors as a separate data set.
- To do this, first load the “plyr” package. Go to Packages & Data, Package Manager and load “plyr”. Then type (in the lower window):

```
SumStats <- ddply(warpbreaks,
  c("wool", "tension") , summarise,
  N = length(breaks),
  mean = mean(breaks),
  sd = sd(breaks),
  se = sd / sqrt(N) )
```

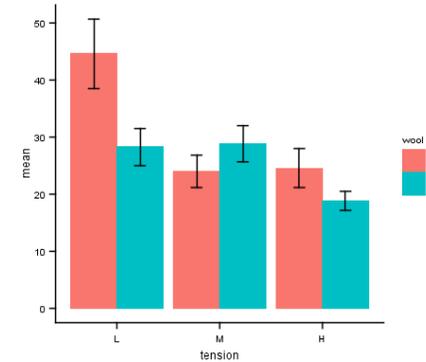
40

Example 4: Bar chart (two factors)

- Go to Plots > Plot Builder.
- NOTE: We are now using the SumStats data!
- **To create the bars:** Under Geometric Elements, choose bar. X = tension, Y = mean, Fill = wool. Position (top left) choose dodge. Instead of bin, choose identity.
- **To add the error bars:** Under Geometric Elements, choose error bar. X = tension, Ymin = mean - se, Ymax = mean + se, width = 0.2, Group = wool. Position (top left) choose dodge, width = 0.9.
- **To change the “theme”:** Under Other, choose classic. You can also change the font size here.

41

Example 4: Bar chart (two factors)



42

Example 4: Two way ANOVA

- Go to Analysis > Linear Model. Outcome = breaks, As Factor: wool, tension. From the Model Builder, Model should include wool, tension, wool:tension (the interaction term). Or equivalently, just wool*tension.
- Under Options, for Tables Type = III. Unselect Summary Table.
- Under Post-Hoc, select both wool and tension. This will give us pairwise comparisons on the main effects.
- NOTE1: Unfortunately no option to get post-hoc comparisons on interaction.
- NOTE2: I really like the “lsmeans” package for running pairwise comparisons, but that is beyond the scope of today’s workshop.

43

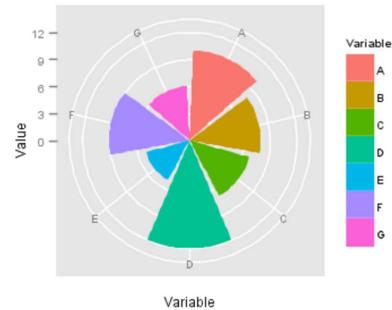
Example 5: Rose Chart

- A toy data set called RoseData1.csv is available from my website.
- This is not a built-in R data set, so it needs to be imported
- We will cover:
 - Rose Chart
 - Making Templates

44

Example 5: Rose Chart

- From Geometric Elements, choose bar. X = Variable, Y= value, Fill = Variable. Instead of bin, choose identity.
- From Coordinates, choose polar.



45

Example 5: Making Templates

- From the Plot Builder window, go to Tools and select Make Template. Choose an appropriate name.
- To save the template, right click and choose Export Template.
- To use the template again, from the Plot Builder window, go to File > Import Template. It will appear under the Templates tab.

46

Saving Work

- There are several options for saving work.
- If you just want to save a “final” graphic (say for inclusion in a paper), then use File > Save As from the graph window.
- If you want to save a graph for further editing, then use File > Save from the Plot Builder window.
- To save commands and/or output, use File > Save from the Console window. You have the option to save (1) Complete Output, (2) Commands or (3) Results.

47