



Jumpstart: Knitr & RMarkdown

NYC Data Science Academy



Outline

- Intro to Knitr and RMarkdown
- Prerequisite
- Markdown Basics
- R Code Chunks
- Plots
- Example



Intro to Knitr and RMarkdown

- R Markdown is a file format used by R to generate dynamic documents that can contain embedded R code and plots.
- Text is written using the markdown format.
 - Markdown formatting is also used by Github and Wordpress.
- Knitr can then be used to output the R markdown document in multiple formats, such as HTML or PDF.



Prerequisite

- Install the following packages:

```
install.packages('dplyr')  
install.packages('ggplot2')  
install.packages('rmarkdown')  
install.packages('knitr')  
install.packages('data.table')
```

- Open the `example.Rmd` file in Rstudio.



Rendering Output

- RStudio: click the "Knit" button (**OSX: Command+Shift+K, Win: Ctrl+Shift+K**)
- Console: `rmarkdown::render` function

```
rmarkdown::render("example.Rmd")
```

- Aside: why is `rmarkdown` followed by the double colon sign?
- The formal name of it is called NAMESPACE. **TL;DR**
- Try the following commands below:

```
library(plyr)  
library(dplyr)  
?summarise
```



Markdown Basics - Emphasis

- Emphasis could be either italic, bold or both.
- It also works on Slack!

```
*italic*    **bold**
```

```
_italic_    __bold__
```

I am italic **I am bold**

```
***italic and bold***
```

italic and bold



Headers

- In markdown, we have headers from 1 to 6 (largest to smallest), depending on the number of # signs you put in front of the text.

```
# Header 1  
## Header 2  
### Header 3
```

Header 1

Header 2

Header 3



Lists

- Bullet lists

- A bullet list is a list of items that begins with a bullet (*, +, or -).

- - Item 1
 - Item 2
 - Item 2a
 - Item 2b

- Ordered lists

- Ordered lists work just like bulleted lists except that the items begin with enumerators rather than bullets.

- - 1. Item 1
 - 2. Item 2
 - 3. Item 3
 - + Item 3a
 - + Item 3b

- **The nested list must be indented four spaces**



Links

- A plain http address or add a link to a phrase.
- **Inline Link:**

```
[The link of your blog post](http://blog.nycdatascience.com/)
```

- The link of your blog post
 - Note that there is no space between the closing bracket and opening parenthesis.
- **Reference Link:**

```
[homepage]: http://nycdatascience.com/  
You can visit our [homepage]
```

You can visit our [homepage](http://nycdatascience.com/)

- For more details on using R Markdown see <http://rmarkdown.rstudio.com>.



R Code Chunks

- We wrap regular R code within three backquotes and assign a name to the chunk, cars in this example. Then we can easily navigate through different code chunks.
- Makes it easy to group related code together so it can be evaluated at the same time.

```
```${r cars, echo=TRUE}  
summary(cars)
```
```

| ## | speed | dist |
|----|--------------|----------------|
| ## | Min. : 4.0 | Min. : 2.00 |
| ## | 1st Qu.:12.0 | 1st Qu.: 26.00 |
| ## | Median :15.0 | Median : 36.00 |
| ## | Mean :15.4 | Mean : 42.98 |
| ## | 3rd Qu.:19.0 | 3rd Qu.: 56.00 |
| ## | Max. :25.0 | Max. :120.00 |



Basic Chunk Options

- `echo(TRUE)`: whether to include R source code in the output file
- `eval(TRUE)`: whether to evaluate the code chunk
- `cache(TRUE)`: if nothing changed in the code chunk from a previous run then it will use the previous result as the output. Good for loading large dataset
- `warning(TRUE)`: whether to preserve warnings in the output
- `message(TRUE)`: whether to preserve message when load the library in the output



Plots

- RMarkdown has its own standalone environment, including working directory.
- It will render the file from the beginning to the end so you need to load the library in the rmarkdown file no matter what you did in the current R session.

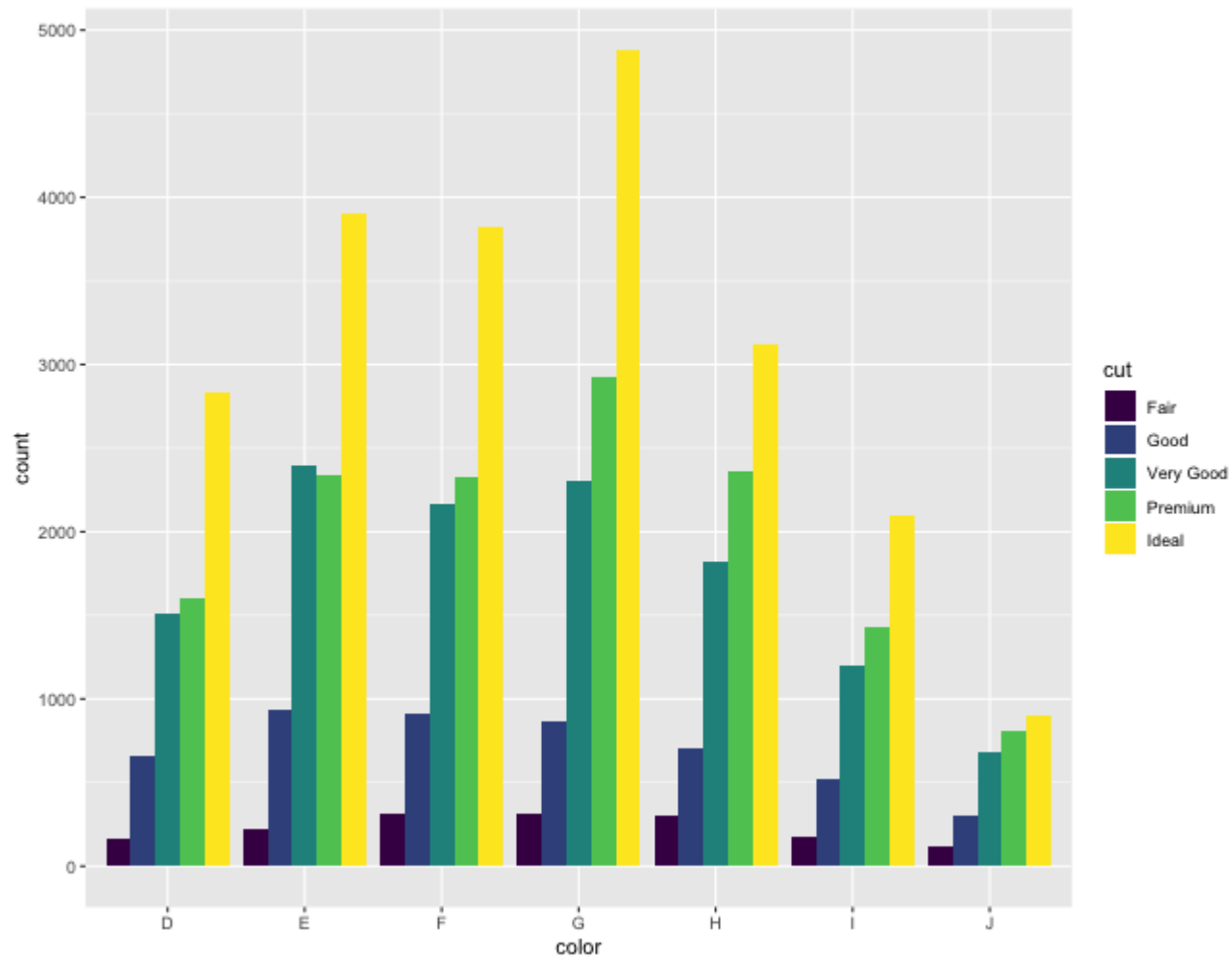
```
```{r echo=TRUE, fig.height=7, fig.width=9, dev='png'}  

library(ggplot2)

ggplot(data = diamonds, aes(x = color)) +
geom_bar(aes(fill = cut), position = "dodge")
```
```



Plots





Plots

- `dev('png')`: figure format(png, jpeg, tiff, svg, ...)
- `fig.path('./figure/')`: figure path (save the image to the figure subdirectory)
- `fig.width(7)`: figure width in inches
- `fig.height(7)`: figure height in inches



Example

- NYC's Department of Health and Mental Hygiene (DOHMH) conducts unannounced inspections of restaurants at least once a year to check food handling, food temperature, personal hygiene and vermin control. Since 2010, NYC restaurants have to prominently post their Grade (e.g. A/B/C) which empowers diners with decision-making information and incentivizes establishments to improve their hygiene.
- Download the csv file from [here](#) and move it to your rmarkdown folder.
- Rename the file to `data.csv`.



Example

- New R code chunk reads in the dataset and produces output related to the dataset.
- It is a good example to set `cache=TRUE` here since you don't want to read the whole csv file everytime you knit the rmarkdown file.
- Always use relative path to read your data instead of `setwd()` to make your project reproducible.

```
```{r load_data, cache=TRUE, warning=FALSE}  
library(data.table)
raw.df <- fread("./data.csv", stringsAsFactors = F)
raw.df <- as.data.frame(raw.df)
str(raw.df)
```
```