Reproducible Reporting Using R Markdown

Case

- 37 y/o M informatician with PMH of email overload disorder
- Request from informatics staff:

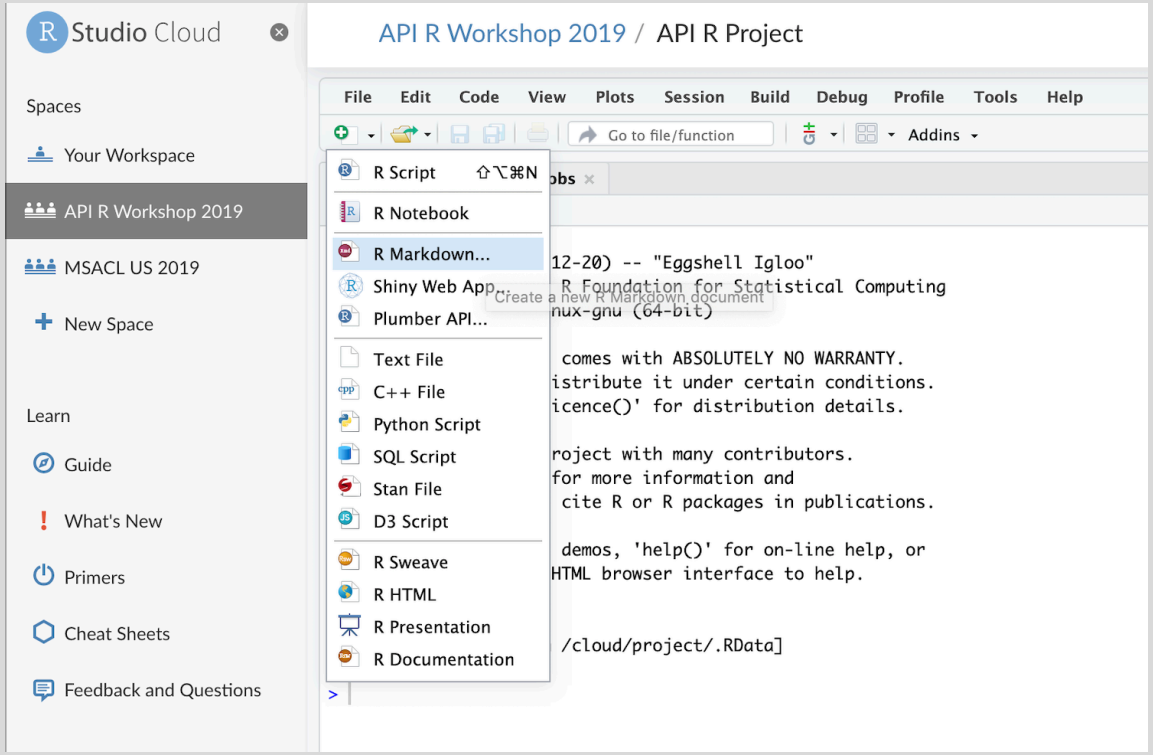
  “Please provide detailed data from your 2 year old analysis of total departmental effort spent performing test cancellations for a SBAR calling out the need to invest effort in duplicate checking rules for the ongoing EHR implementation project”
- Multiple poorly commented scripts titled analysis_1.R, analysis_2.R, etc.

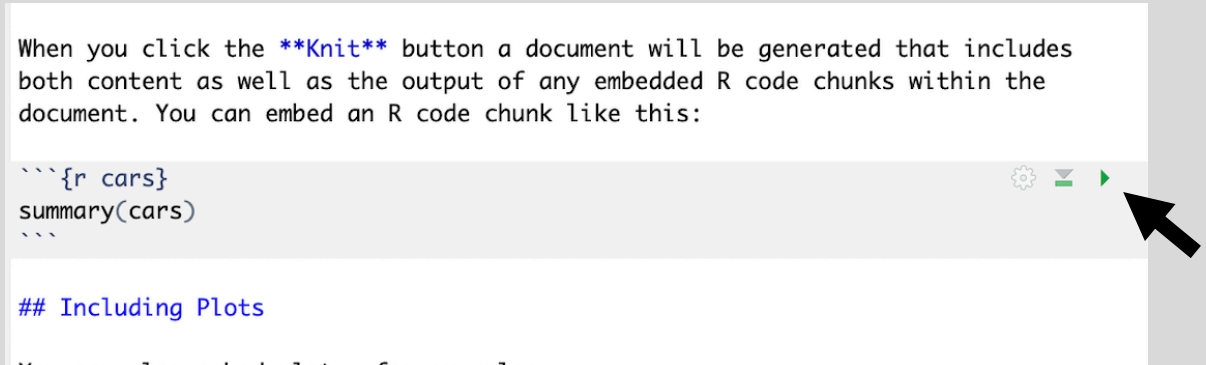
- Consider the above scenario, but with someone else performing the original analysis
- Would it be less work to start from scratch and rewrite the analysis?

Why integrate your analysis and documentation in one place?

1. It will be easier for you to understand what you did when you come back to it 1 year later.
2. It will be easier for anyone to understand what you did.
3. Communication to yourself and others will help you think more about your code and improve it.

Your Turn





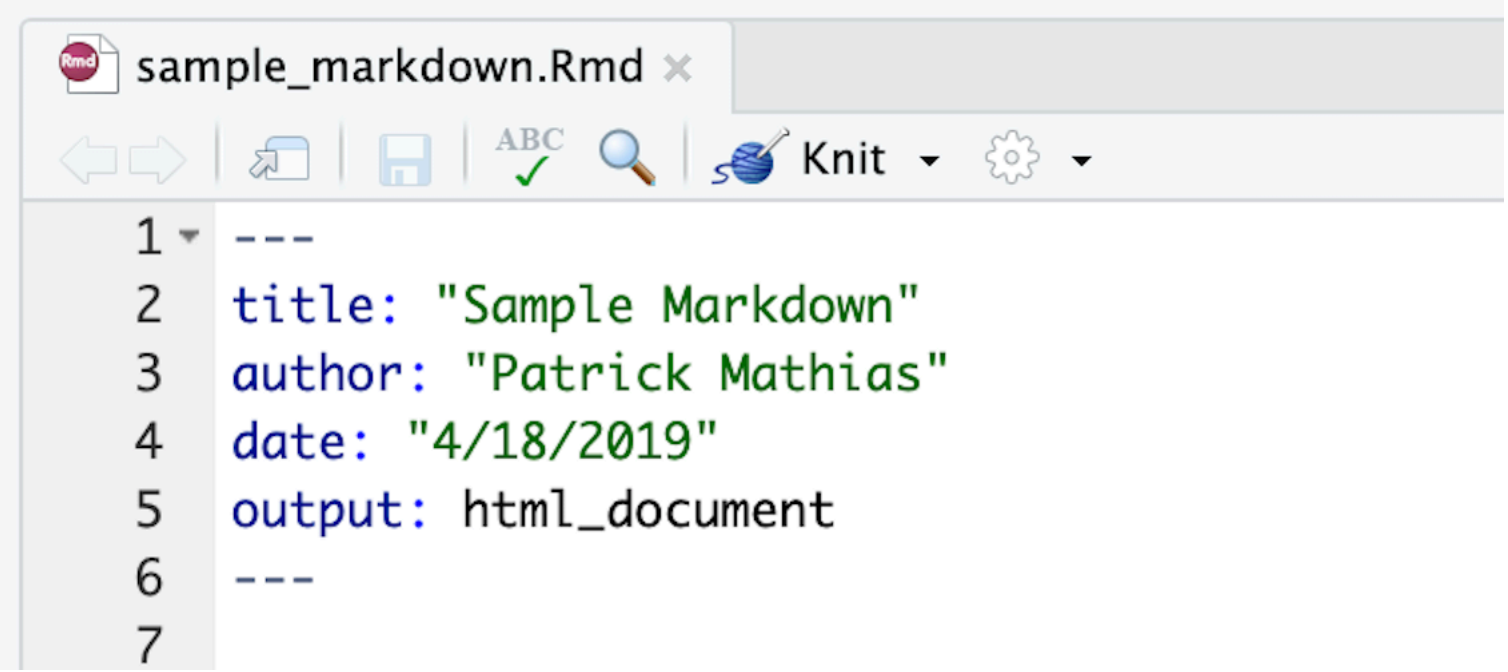
- Open a new R Markdown document within RStudio Cloud
- Enter a Title and Author and leave the output format as HTML
- Explore the document and buttons within the document
- "Knit" the document and save the file

03:00

Header

- Separated from document by ---
- Includes metadata about document
- YAML format

  Field name: Data



Text

```
11
12 ## R Markdown
13
14 This is an R Markdown document. Markdown is a simple formatting syntax for
   authoring HTML, PDF, and MS Word documents. For more details on using R
   Markdown see <http://rmarkdown.rstudio.com>.
15
16 When you click the **Knit** button a document will be generated that
   includes both content as well as the output of any embedded R code chunks
```

- Uses markdown syntax
- # for headers
  - # for level 1
  - ## for level 2

- 1 asterisk for *italics* (*italics*)
- 2 asterisks for **bold** (**bold**)
- Hyphens for bullet points

Code chunks


Open/close
Language
Name of chunk
Run chunk
17
18 ```{r cars}
19 summary(cars)
20 ```
21
16 When you click the **Knit** button a document
includes both content as well as the output of
within the document. You can embed an R code c
17
18 ```{r cars}
19 summary(cars)
20 ```
21
22 ## Including Plots
23
24 Yo s, for example:
25
26 ``` LSE}
27 pl
28 ```
29
30 No ALSE` parameter was add
Sample Markdown
Chunk 1: setup
R Markdown
Chunk 2: cars
Including Plots
Chunk 3: pressure
2:1
Sample Markdown
Console
Terminal
Jobs
/cloud/project/
R version 3.5.2 (2018-12-20) -- "Eggshell Igloo"

Your Turn

- Insert a code chunk into white space within your open R Markdown document using:
  - Windows: CTRL+ALT+i
  - Mac: COMMAND+OPTION+i
- Add the following to your new code chunk:

  ```
  mean(c(10, 20, 30))
  ```

- Execute using shortcuts:
  - Windows: CTRL+SHIFT+ENTER
  - Mac: COMMMAND+SHIFT+ENTER
- Include the following within any text area (white space) and knit:

  ```
  `r mean(c(10, 20, 30))`
  ```

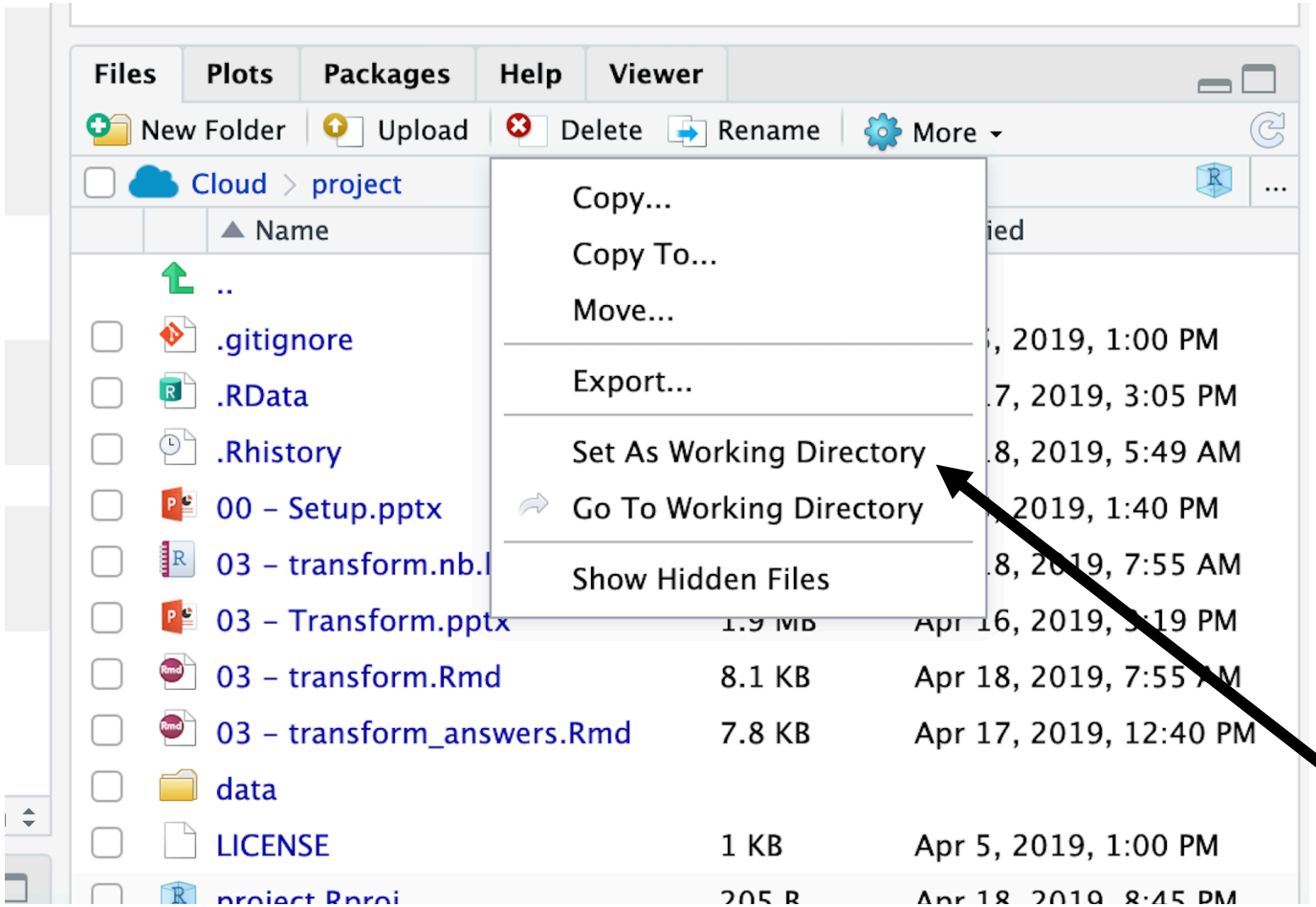
03:00

Working with R Markdown for this course

- Each lesson has an R Markdown file
  - Executable examples
  - Exercises
- Files used as “notebooks”: can document, execute, and iterate
- Best practice: work from notebook rather than console

Open “02 – Report.Rmd” and run the setup chunk.

Where am I? Your working directory



- Navigate folder and file structure on your computer from Rstudio
- getwd() function will tell you which folder you're in
- setwd() to set a new working directory

Can navigate folders and set working directory with this menu

Reading comma separated or tab delimited files


```
read_csv("data/test_menu.csv")
```

File type (csv, tsv, delim for non-standard delimiters)

File path (if file in a folder within working directory)

File name

Reading Excel files

File name

```
read_excel("data/excel_file.xlsx",
           sheet = 1)
```

Specify sheet by number or "name"

Can also extract arbitrary rows and columns using "range = " argument

Your Turn

Find and import the orders data set and view the data

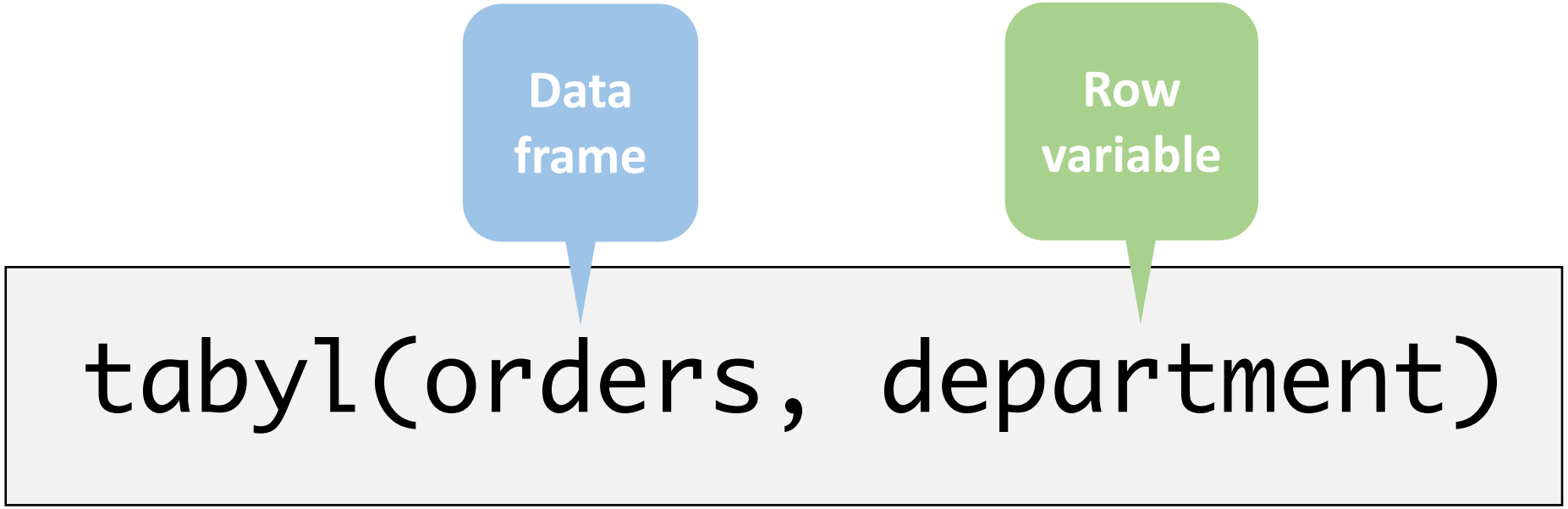
order_id	patient_id	description	proc_code	order_class_c_descr	lab_status_c	lab_status_c_descr	order_status_c	order_status_c_descr
19766	511388	PROTHROMBIN TIME	PRO	Normal			4	Canceled
88444	511388	BASIC METABOLIC PANEL	BMP	Normal			4	Canceled
40477	508061	THYROID STIMULATING HORMONE	TSH	Normal	3	Final result	5	Completed
97641	508061	T4, FREE	T4FR	Normal	3	Final result	5	Completed
99868	505646	COMPREHENSIVE METABOLIC PANEL	COMP	Normal	3	Final result	5	Completed

03:00

View a quick summary

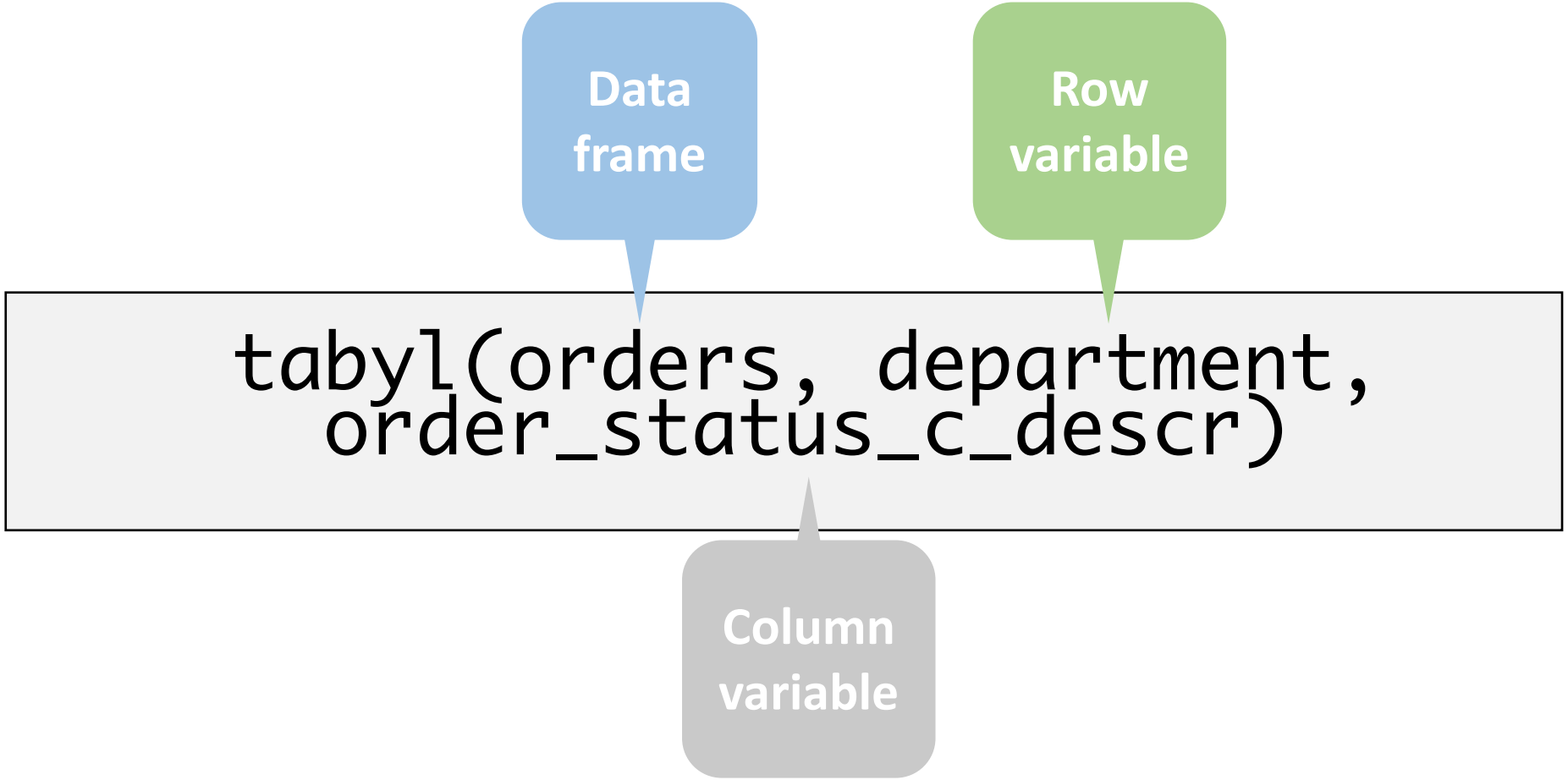
- summary( ) function outputs quick statistical summaries for numerical and timestamp fields
- Provides limited data (only a count) for character fields
- Will provide counts of different categories for factor fields (categorical)

Create a one variable table



Two variable tables can help answer simple questions quickly

Which clinic cancelled the highest number of lab orders?



Your Turn

Which 3 departments ordered the highest number of labs using “Provider Preference Lists”?

03:00

Outline

- Create a new markdown document
- R Markdown structure
- Open up lesson Rmd
- Demo loading file from Rmd – cover working directory
- Exercise: load orders file
- Plots
- Knit other file types