

# Doing Science : from Start to Finish (day 2)

Dr. Félix E. Rivera-Mariani

<https://github.com/friveramariani/DoingScienceWorkshops>

# Today's goals

---

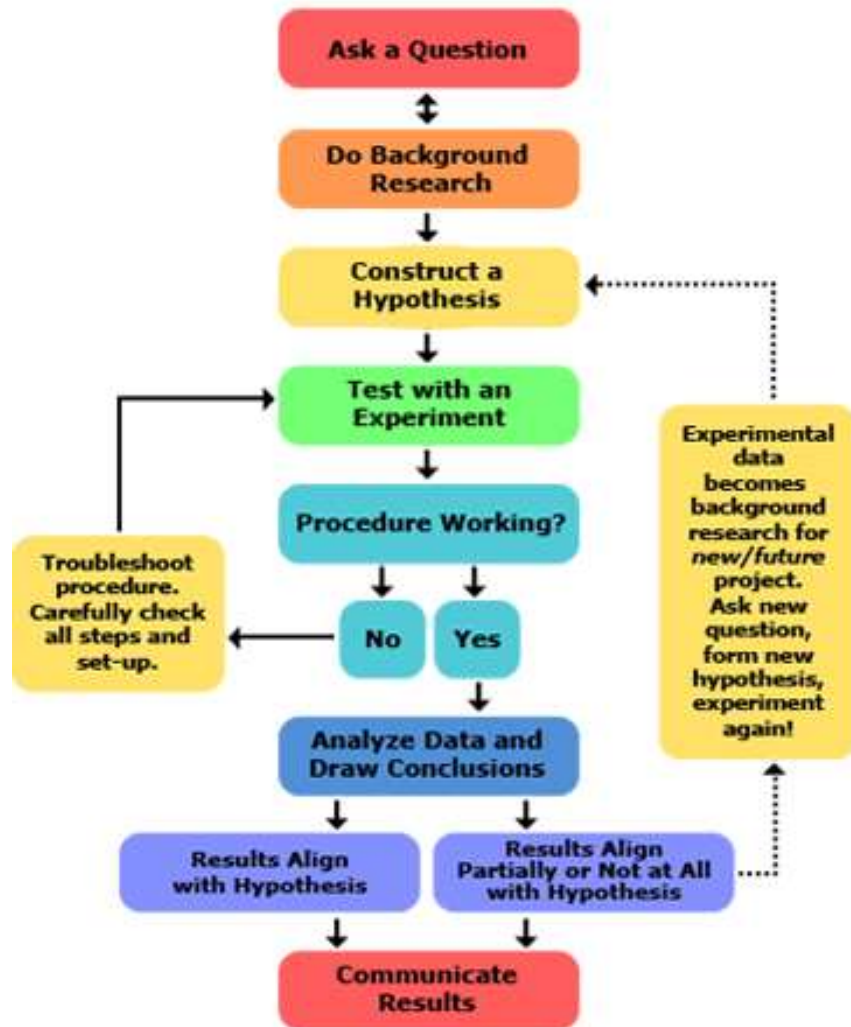
- Go through all stages of the scientific method

# Today's goals

---

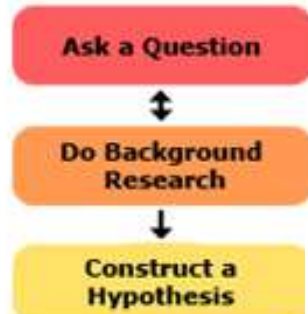
- Go through all stages of the scientific method
- Design an experiment to answer a question for which we don't know much about topic

# The Scientific Method



- Two directions between question and background research
- Epicycles with the experiment
- Epicycle between results and hypothesis

# The Scientific Method: *important features*



- *From **Question** to **Hypothesis***

# The Scientific Method: *important features*



- **From *Question* to *Hypothesis***
  - *Specific*
  - *Testable*
  - *Falsifiable (applies to hypotheses)*

## Scientific Method



The Scientific Method: *a good question, hypothesis, or purpose*

# The Scientific Method: *question's criteria*

---

## Scientific Method

Ask a Question



# The Scientific Method: *questions' criteria*

## Scientific Method

Ask a Question

- To determine how long will it take me to go through each room of the Thomas Armstrong high school without exiting from the same door that I entered.

Is this question specific? **Yes**

Can it be tested? **Maybe**

Can we develop a hypothesis or should we use a purpose?

- **We lack information for a hypothesis:**
  - Do all rooms have two doors
  - We do not know how long it takes to go through one room.
  - We do not know how many rooms are in the schools.
- **We must proceed with a purpose: To determine how long will it take me to go....**

# The Scientific Method: *questions' criteria*

## Scientific Method

Ask a Question

- What are the most frequent words in the students expectations' about the workshop *"Doing Science: from Start to Finish"*?

Is this question specific? **Yes**

Can it be tested? **Yes**

Can we develop a hypothesis or should we use a purpose? **Yes**

- **We know what is the workshop about**
- **We know keywords related to the workshop**
- *Then, we can proceed with a hypothesis*

# The Scientific Method: *developing a hypothesis*

---

## **Scientific Method**

**Construct a  
Hypothesis**

# The Scientific Method: *developing a hypothesis*

---

## Scientific Method

- What are the most frequent words in the students expectations' about the workshop "*Doing Science: from Start to Finish*"?

**Construct a  
Hypothesis**

# The Scientific Method: *developing a hypothesis*

## Scientific Method

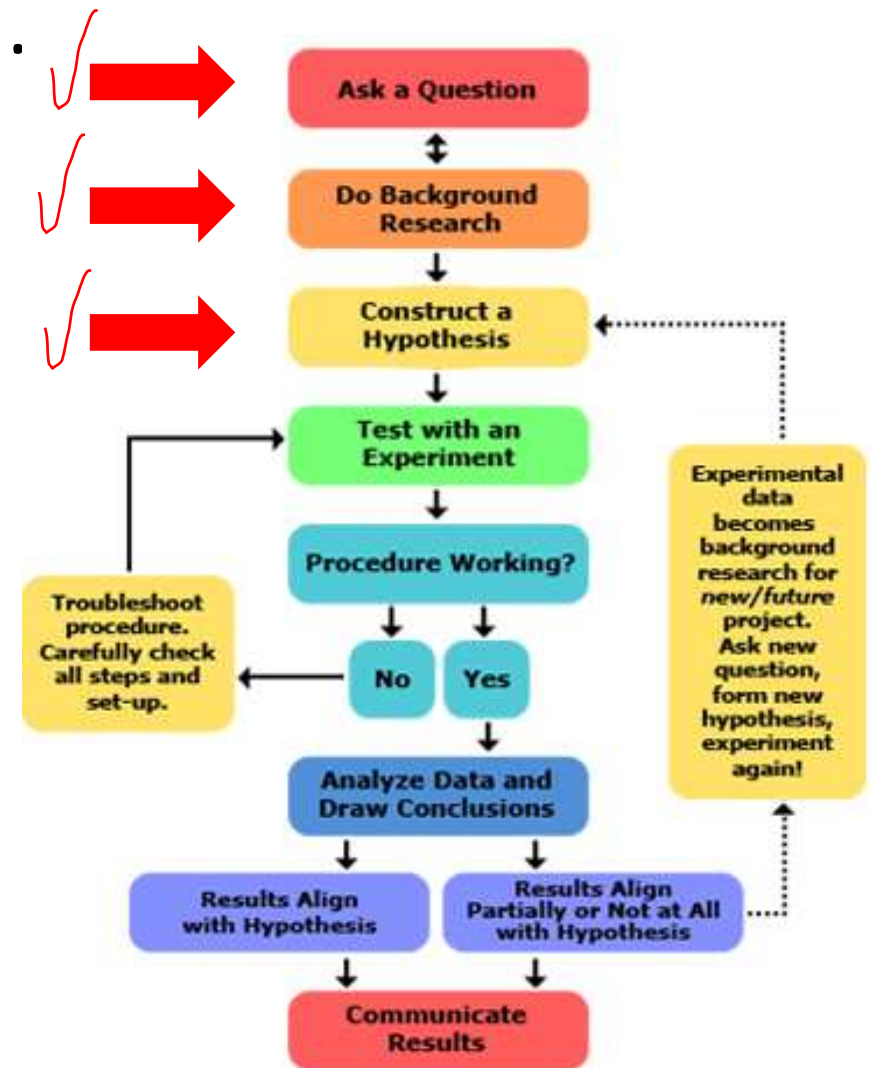
- What are the most frequent words in the students expectations' about the workshop "*Doing Science: from Start to Finish*"?

**Construct a Hypothesis**

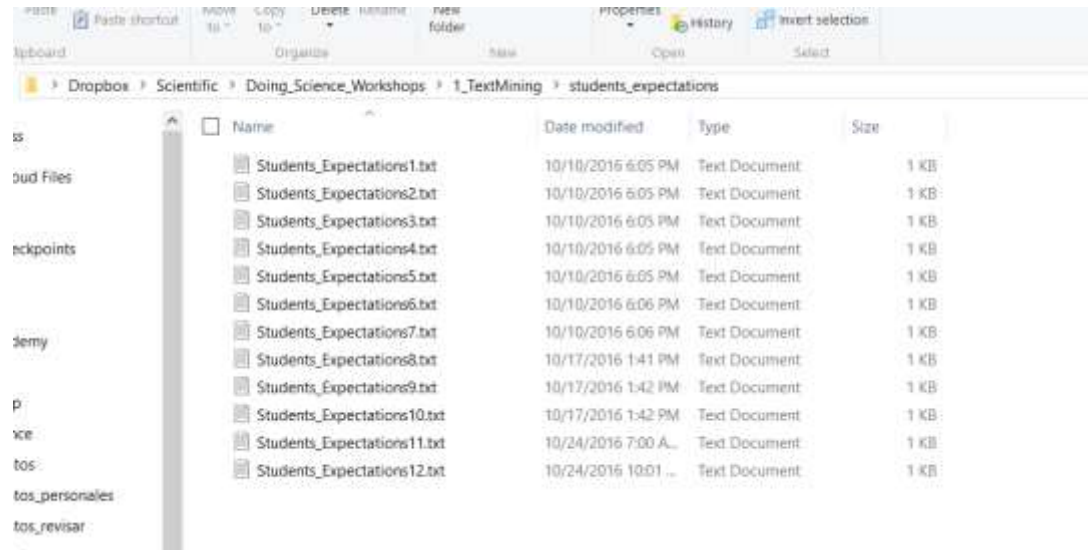
**My hypothesis:** The students expectations' of the Doing Science workshop will contain words that are in the title of the workshop.

My hypothesis: The students expectations' of the Doing Science workshop will contain words that are in the title of the workshop. ✓

Then, what's next?



**My hypothesis:** The students expectations' of the Doing Science workshop will contain words that are in the title of the workshop.



The screenshot shows a file explorer window with the following path: Dropbox > Scientific > Doing\_Science\_Workshops > 1\_TextMining > students\_expectations. The table below lists the files in this directory.

Name	Date modified	Type	Size
Students_Expectations1.txt	10/10/2016 6:05 PM	Text Document	1 KB
Students_Expectations2.txt	10/10/2016 6:05 PM	Text Document	1 KB
Students_Expectations3.txt	10/10/2016 6:05 PM	Text Document	1 KB
Students_Expectations4.txt	10/10/2016 6:05 PM	Text Document	1 KB
Students_Expectations5.txt	10/10/2016 6:05 PM	Text Document	1 KB
Students_Expectations6.txt	10/10/2016 6:06 PM	Text Document	1 KB
Students_Expectations7.txt	10/10/2016 6:06 PM	Text Document	1 KB
Students_Expectations8.txt	10/17/2016 1:41 PM	Text Document	1 KB
Students_Expectations9.txt	10/17/2016 1:42 PM	Text Document	1 KB
Students_Expectations10.txt	10/17/2016 1:42 PM	Text Document	1 KB
Students_Expectations11.txt	10/24/2016 7:00 A...	Text Document	1 KB
Students_Expectations12.txt	10/24/2016 10:01 ...	Text Document	1 KB

- Collected the data

My hypothesis: The students expectations' of the Doing Science workshop will contain words that are in the title of the workshop.

- Test with an experiment

```
#-----Text file-----
# load packages
library(tm)
library(SnowballC)
library(ggplot2)
library(wordcloud)
library(cluster)
library(fpc)

# load directory
std_exp <- "C:/Users/Felix/Dropbox/Scientific/Doing_Science_workshops/1_TextMining/students_expectations"

# create corpus
std_exp_corp <- Corpus(DirSource(std_exp))

# inspect files
inspect(std_exp_corp)

#-----clean txt file-----
# remove punctuations from text files
std_exp_nopunt <- tm_map(std_exp_corp, removePunctuation)

# remove numbers from text files
std_exp_nonum <- tm_map(std_exp_nopunt, removeNumbers)

# all lowercase
std_exp_lowercase <- tm_map(std_exp_nonum, tolower)

# remove common words in spanish and english words
std_exp_rm_common <- tm_map(std_exp_lowercase, removeWords, stopwords("english"))

# remove white spaces
std_exp_nowhite <- tm_map(std_exp_rm_common, stripWhitespace)

# text into plain texts
std_exp_plain <- tm_map(std_exp_nowhite, PlainTextDocument)

# create documents matrices
std_exp_mtx <- DocumentTermMatrix(std_exp_plain)

# transposed document matrices
std_exp_mtx_t <- TermDocumentMatrix(std_exp_plain)

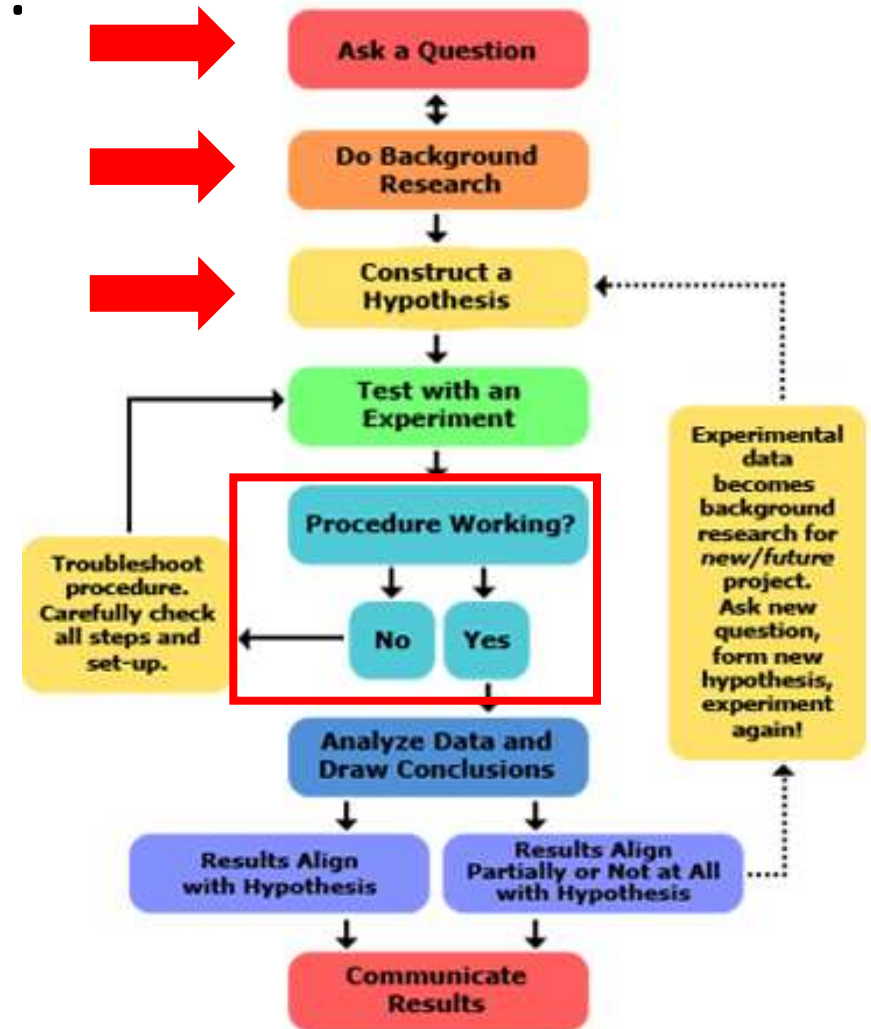
#-----exploratory analysis-----
# create, order, and examine frequencies
freq_stdexp <- colSums(as.matrix(std_exp_mtx_t))
ord_freq_stdexp <- order(freq_stdexp)
freq_stdexp[head(ord_freq_stdexp)]

# create, order, and examine frequencies of w/o sparse terms, spanish terms
dtns_stdnspr <- removeSparseTerms(std_exp_mtx, 0.50)
freq_std_nspr <- colSums(as.matrix(dtns_stdnspr))
ord_freq_std_nspr <- order(freq_std_nspr)
freq_std_nspr[head(ord_freq_std_nspr)]
```

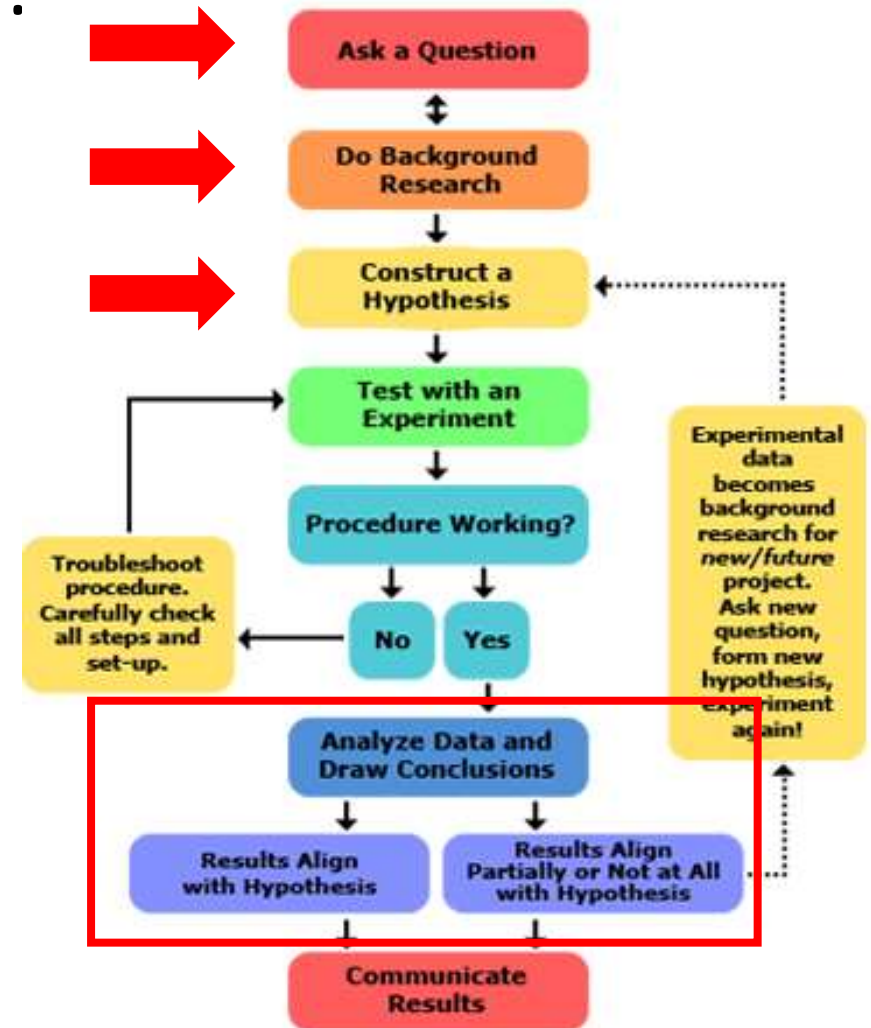
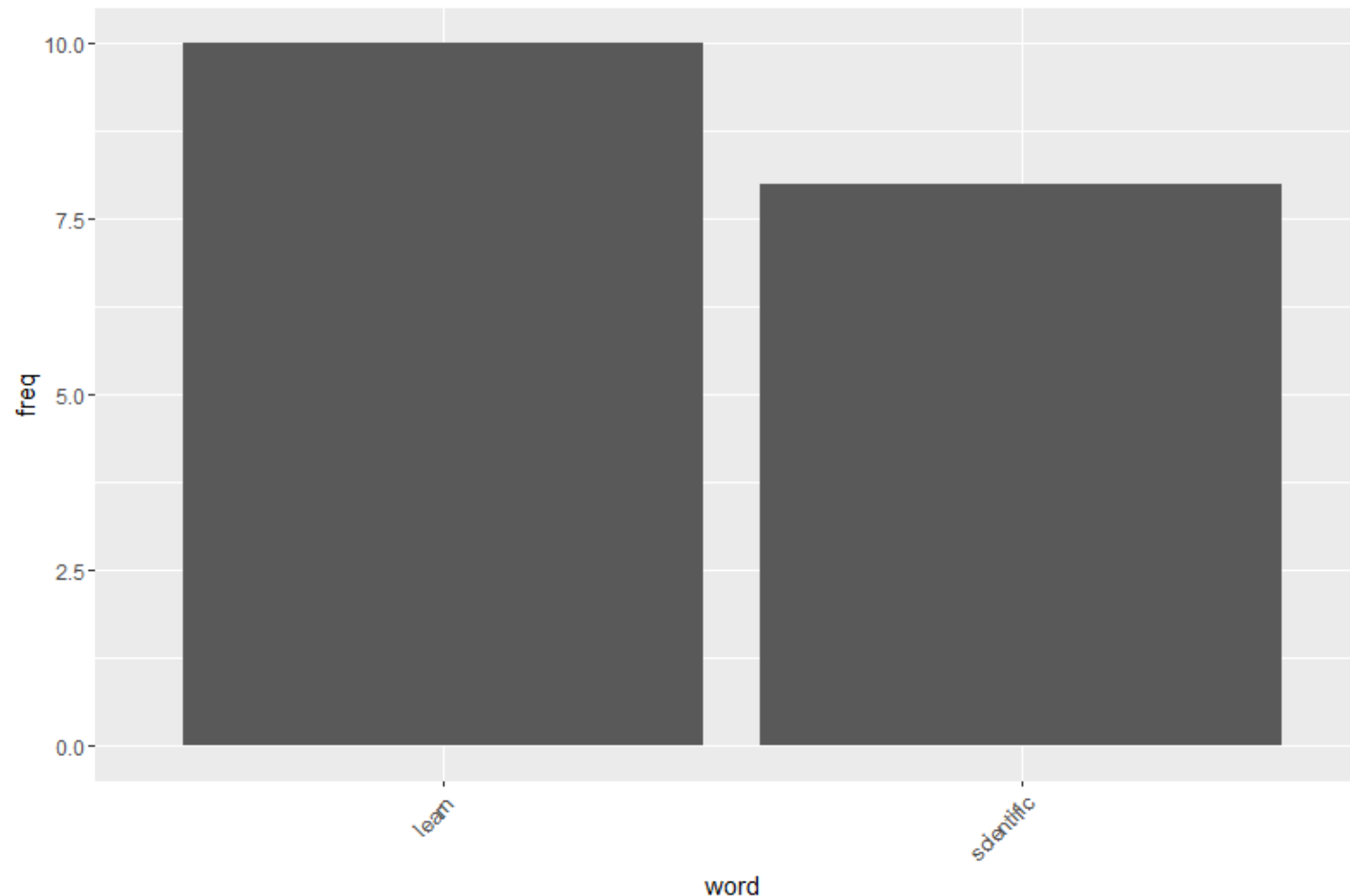


My hypothesis: The students expectations' of the Doing Science workshop will contain words that are in the title of the workshop.

Did the procedure work?



My hypothesis: The students expectations' of the Doing Science workshop will contain words that are in the title of the workshop.



# Homework in groups: re-submit the hypothesis

---

- Re-submit a hypothesis to the question: *What are the most frequent words among the students expectations' of the "Doing Science" workshop?*