Welcome to Week #2!

General Survey Results

Most of programming experience at about 1 year

Wide range of data interests

People are excited about learning more about programming

Most of programming experience at about 1 year

Exercises will have wide range of options for different levels.

Wide range of data interests

Will use “real” datasets, wide variety of topics (feel free to suggest some!)

People are excited about learning more about programming

We will do a lot of collaborative coding in class, different options for different levels.

Week	Topic	Reading
1	• Data, Models, and Information • Elementary statistics: Definitions • Overview of R	OIS 1 (ISL 1)
2	• Elementary statistics: Applications & Plots	OIS 1 (ISL 1)
3	• Introduction to data analysis with R • Review of tabular and graphical displays of data	ITR 1, 2, 5, 6, 7, 12
4	• Random variables: expectation and variance • Joint and conditional probability • Bayes rule	OIS 2
5	• Random variables: distributions (normal, binomial, poisson)	OIS 3

Definitions, basic concepts, R practice

HW and Exam Formats

File name structure: lastname-first-module.ext
(e.g, naiman-jill-assignment1.pdf).

The submission must include:

1) A narrative document as a PDF file (to be read by a human). To preserve the natural flow of the narrative, figures (e.g., screenshots, code snippets) and tables should be embedded into the document near their first mention. Any supplementary files containing R programs or data should be referenced in the text and separately uploaded.

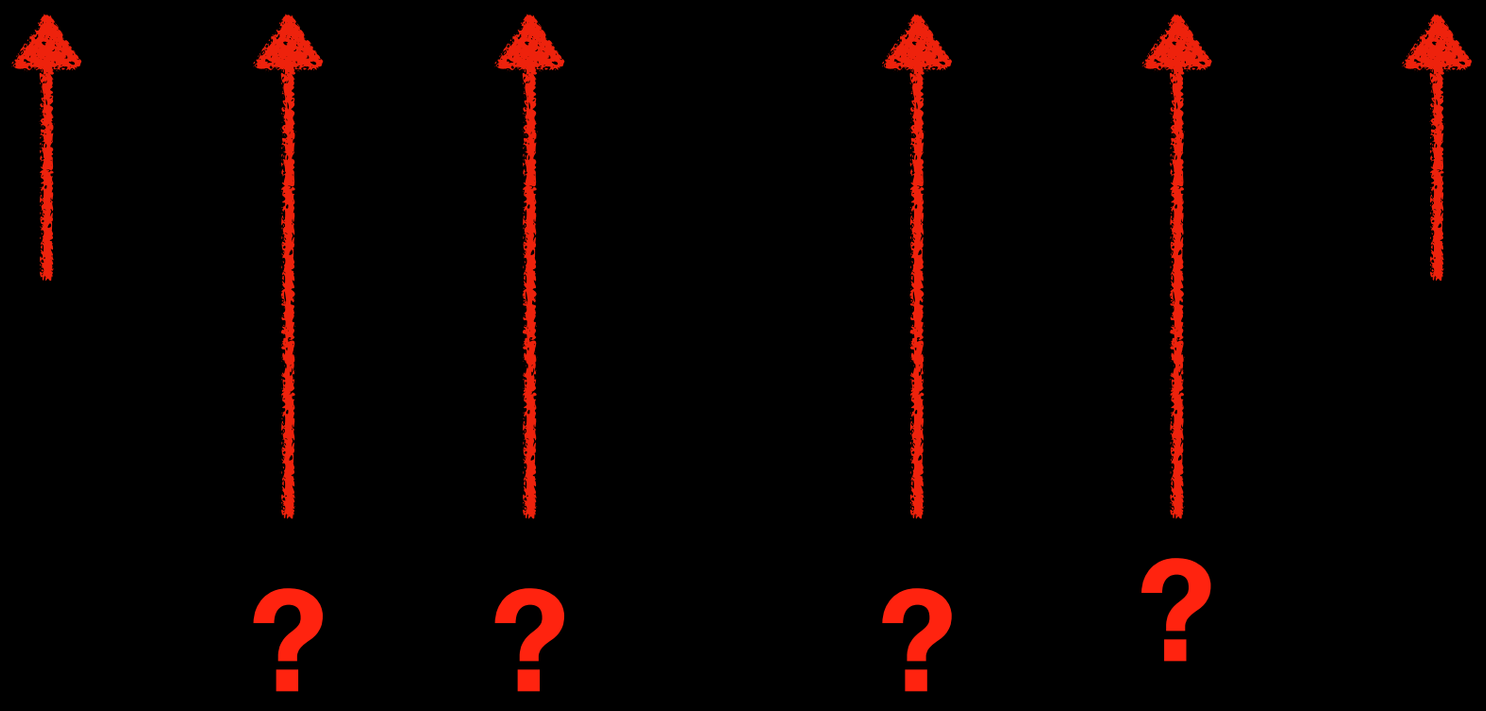
AND

2) All R code as separate files with an .R extension (to be read by a computer).

ALSO: make sure you include any files needed to run your R-script (data files for example)

Last time…

```
summary(after)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  39.00   41.25   44.50   44.21   46.75   50.00
```



Last time…

Summary Statistic Definitions!

Mean (Sample) = sum of all data values divided by number of data points

$$\text{Mean} = \frac{\text{Sum of all values}}{\text{Number of values}}$$

$$\text{Symbolically, } \bar{x} = \frac{\sum x}{n}$$

where $\bar{x}$ (read as 'x bar') is the mean of the set of x values,
$\sum x$ is the sum of all the x values, and
n is the number of x values.

(note - only works with "numerical" data types… more about data types later)

Median = if we order the data from smallest to largest, this is the observation in the middle (splits the data in 2 halves)

First/Third Quartiles = where 25% of the data falls below/above

Standard Deviation = this is the square root of the variance, where the variance is roughly the average distance of data values from the mean

$$\text{Standard Deviation (sample)} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \text{mean})^2}{n-1}}$$

Last time…

Summary Statistic Definitions!

Mean (Sample) = sum of all data values divided by number of data points

$$\text{Mean} = \frac{\text{Sum of all values}}{\text{Number of values}}$$

$$\text{Symbolically, } \bar{x} = \frac{\sum x}{n}$$

where $\bar{x}$ (read as 'x bar') is the mean of the set of x values,
$\sum x$ is the sum of all the x values, and
n is the number of x values.

(note - only works with "numerical" data types… more about data types later)

Median = if we order the data from smallest to largest, this is the observation in the middle (splits the data in 2 halves)

First/Third Quartiles = where 25% of the data falls below/above

Standard Deviation = this is the square root of the variance, where the variance is roughly the average distance of data values from the mean

$$\text{Standard Deviation (sample)} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \text{mean})^2}{n-1}}$$

Last time…

Summary Statistic Definitions!

Mean (Sample) = sum of all data values divided by number of data points

$$\text{Mean} = \frac{\text{Sum of all values}}{\text{Number of values}}$$

$$\text{Symbolically, } \bar{x} = \frac{\sum x}{n}$$

where $\bar{x}$ (read as 'x bar') is the mean of the set of x values,
$\sum x$ is the sum of all the x values, and
n is the number of x values.

(note - only works with "numerical" data types… more about data types later)

Median = if we order the data from smallest to largest, this is the observation in the middle (splits the data in 2 halves)

First/Third Quartiles = where 25% of the data falls below/above

Standard Deviation = this is the square root of the variance, where the variance is roughly the average distance of data values from the mean

$$\text{Standard Deviation (sample)} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \text{mean})^2}{n-1}}$$

Lets play with *your* survey data!

IS542 Spring 2020 Poll

This is a quick, informal poll to gauge skill with statistics & programming to guide development of the course.

* Required

Name *

Your answer

How would you gauge your familiarity with Statistics? *

	1	2	3	4	5	
Novice	○	○	○	○	○	Expert

How would you gauge your familiarity with programming?

Lets play with *your* survey data!

IS542 Spring 2020 Poll

This is a quick, informal poll to gauge skill with statistics & programming to guide development of the course.

* Required

Name *

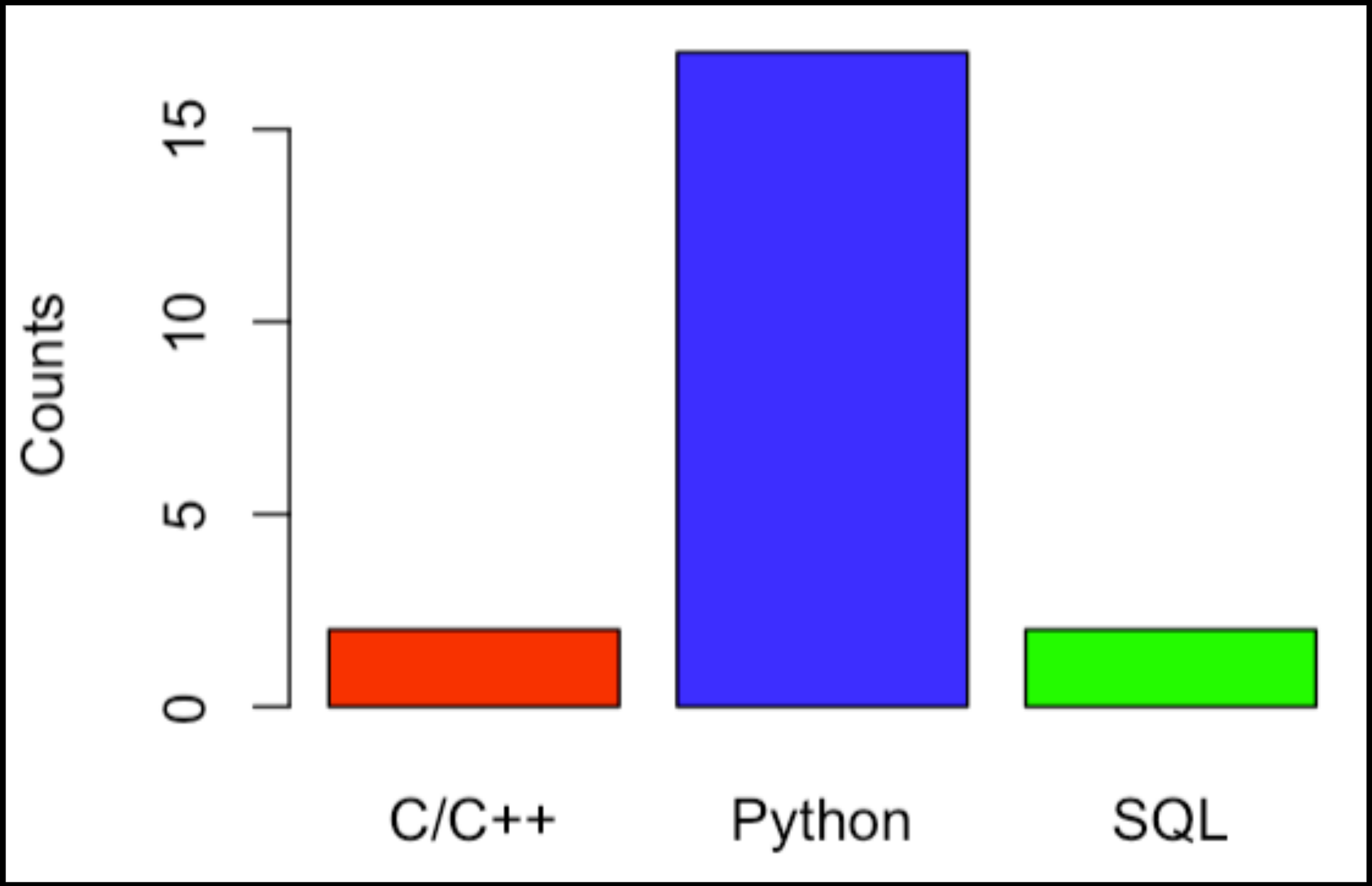
Your answer

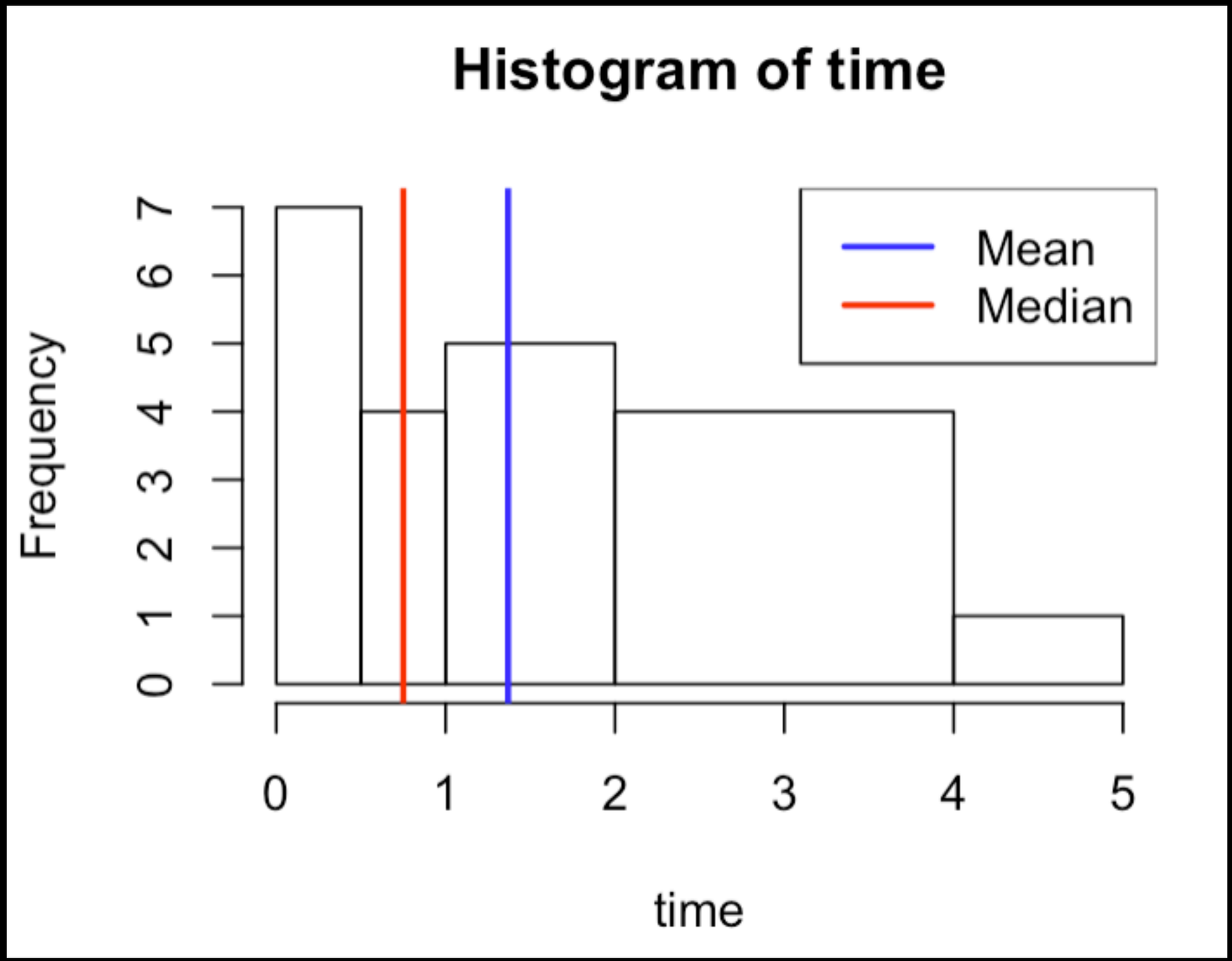
To R!

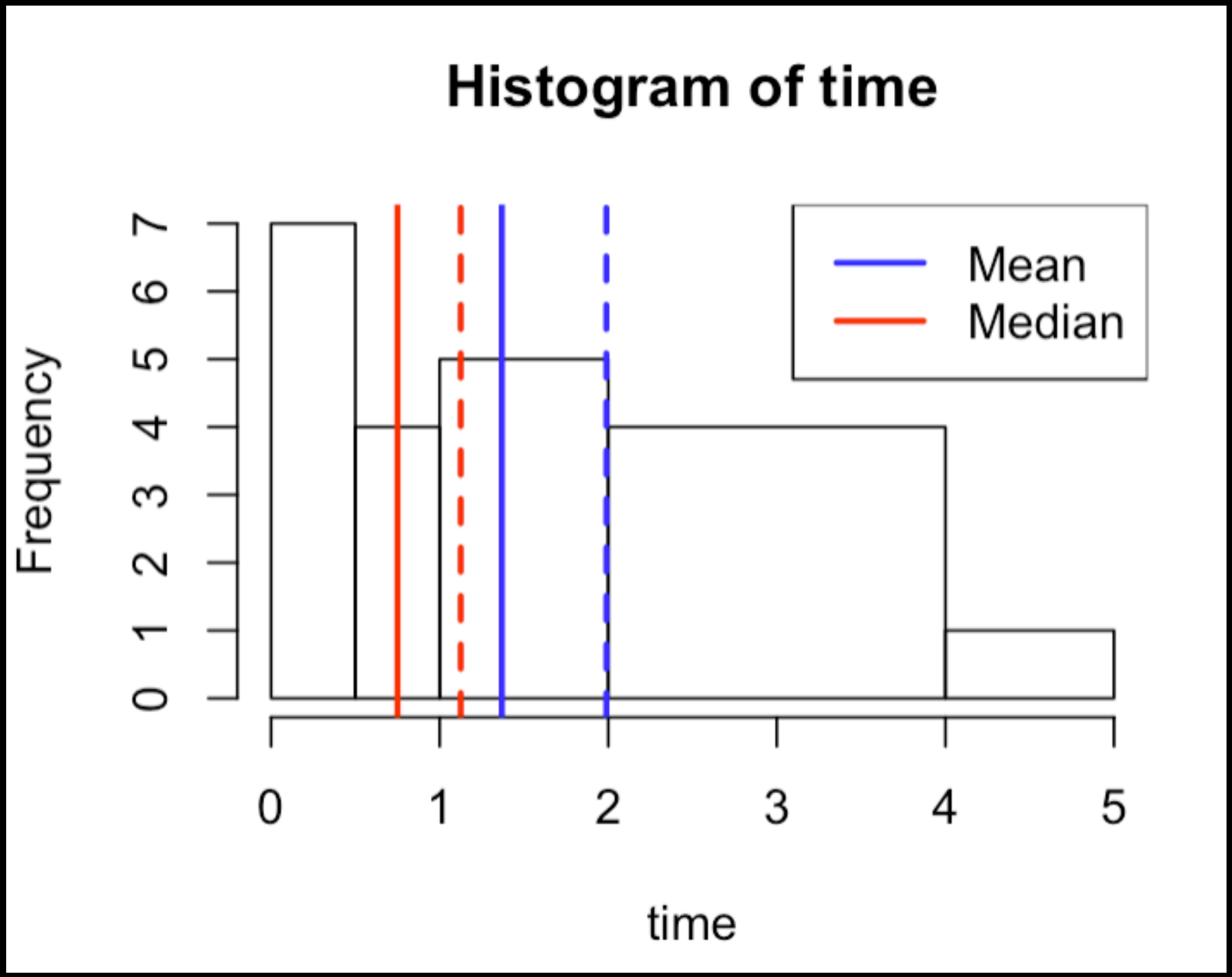
How would you gauge your familiarity with Statistics? *

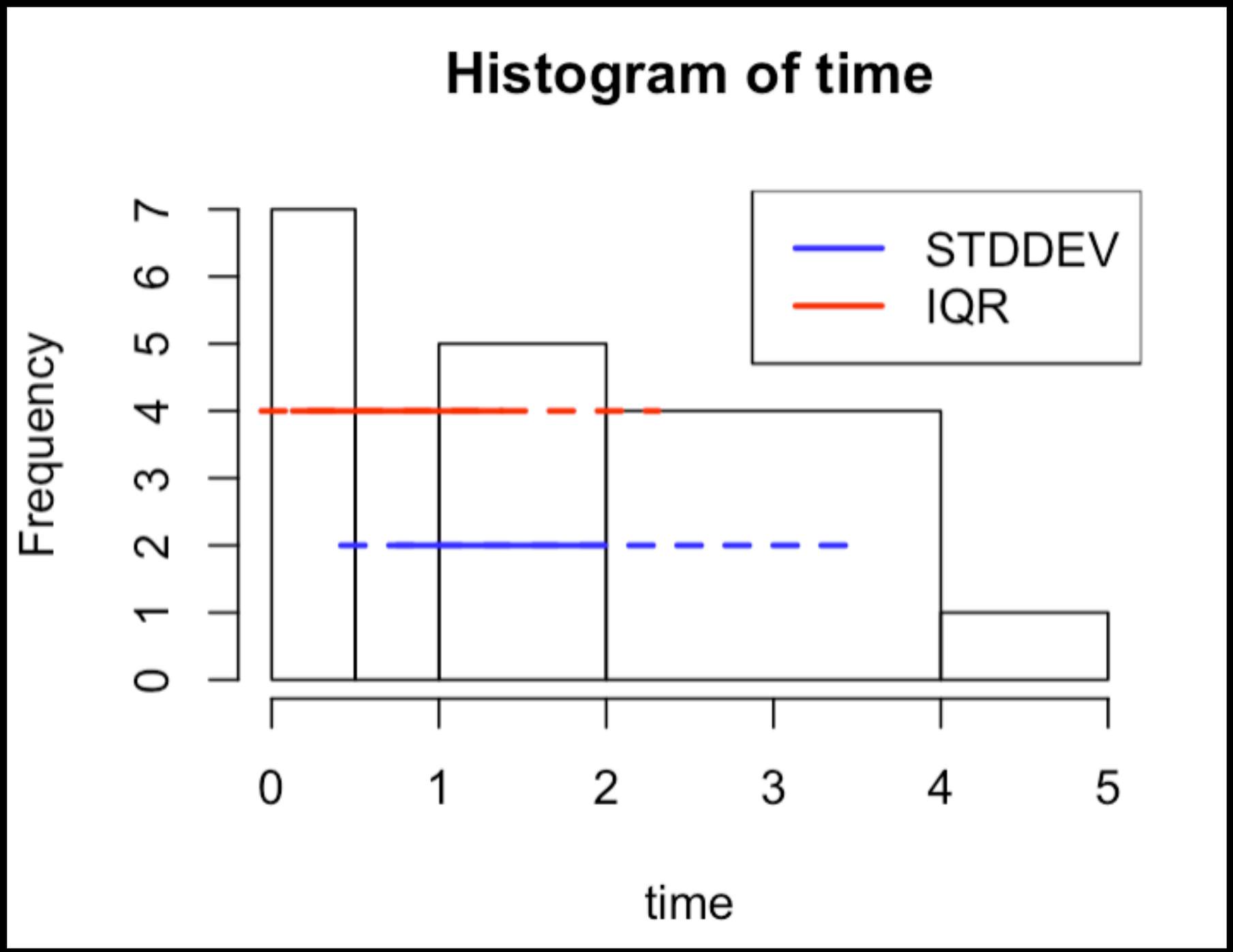
	1	2	3	4	5	
Novice	○	○	○	○	○	Expert

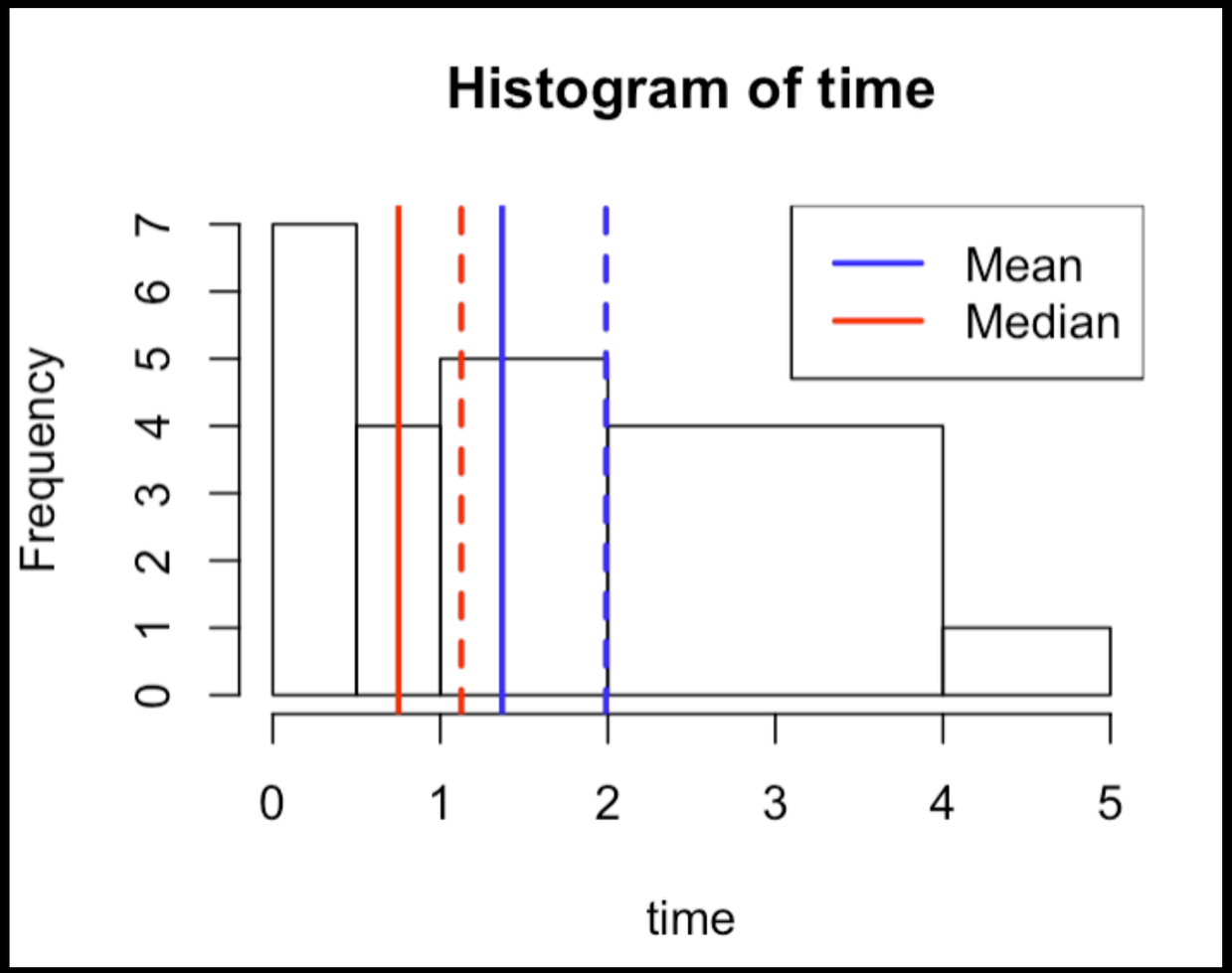
How would you gauge your familiarity with programming?


Counts
0
5
10
15
C/C++
Python
SQL

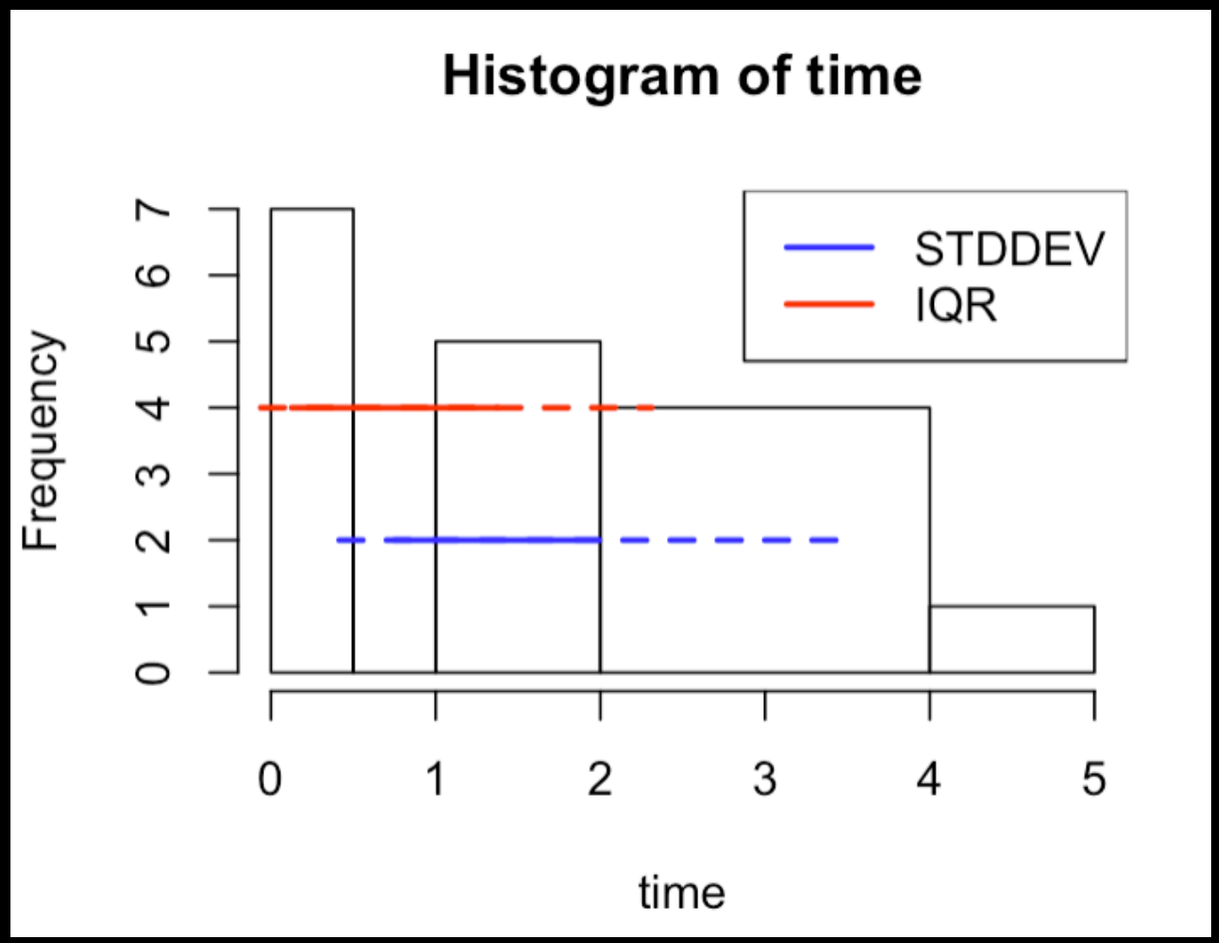

Histogram of time
Mean
Median
Frequency
0
1
2
3
4
5
6
7
0
1
2
3
4
5
time

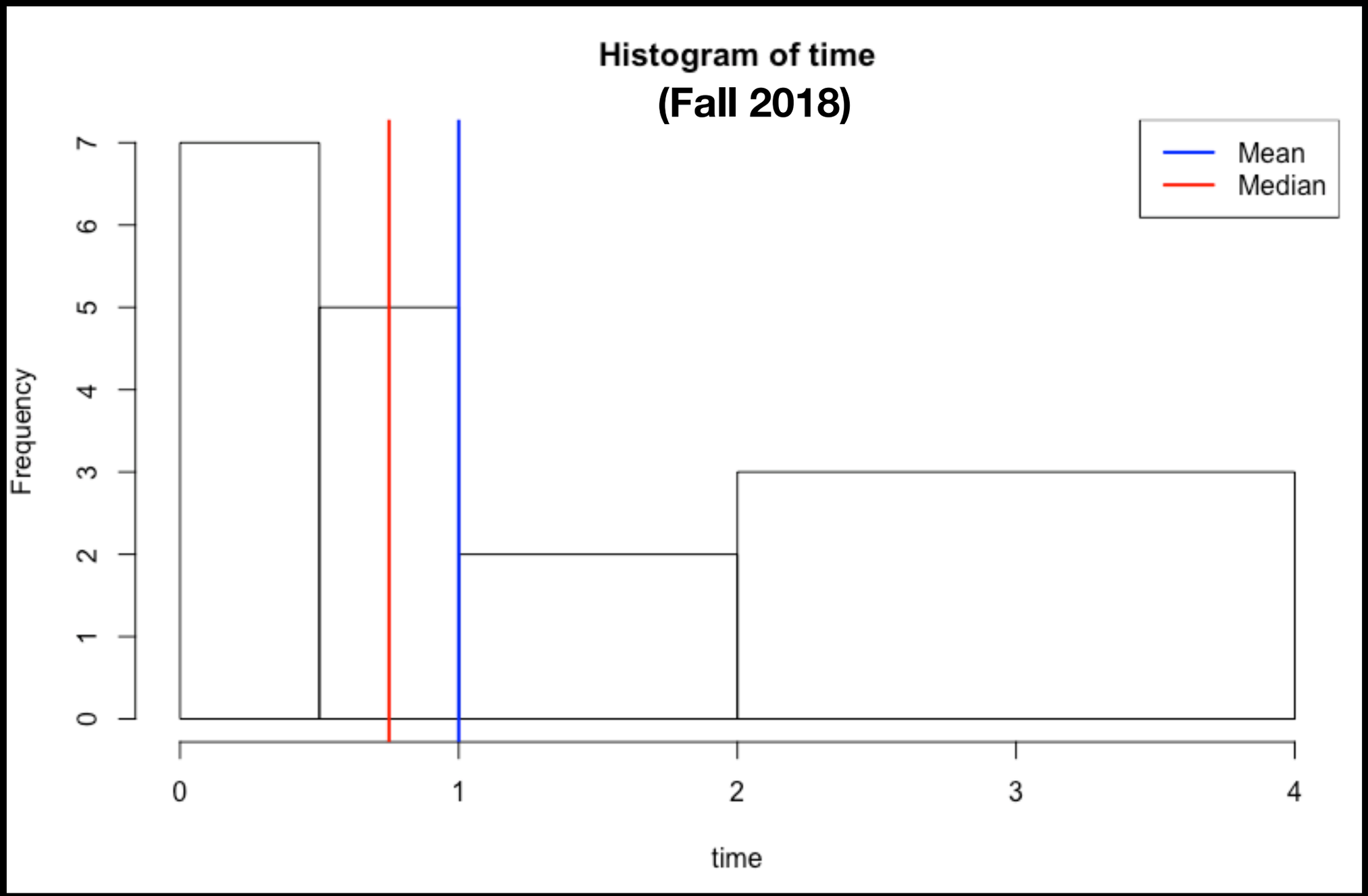

Histogram of time
Mean
Median
Frequency
0
1
2
3
4
5
6
7
0
1
2
3
4
5
time

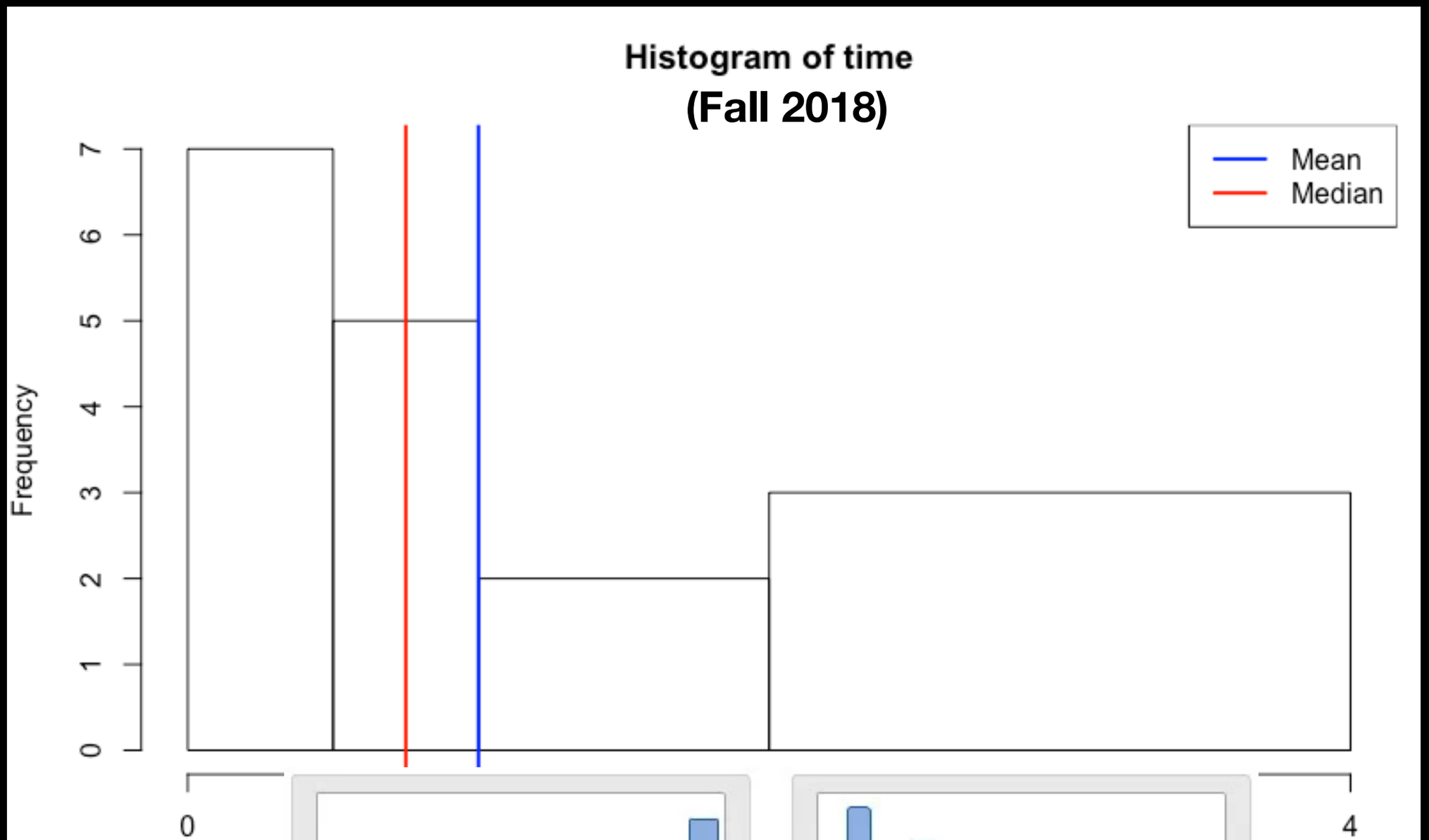

Histogram of time
STDDEV
IQR
Frequency
0
1
2
3
4
5
6
7
0
1
2
3
4
5
time



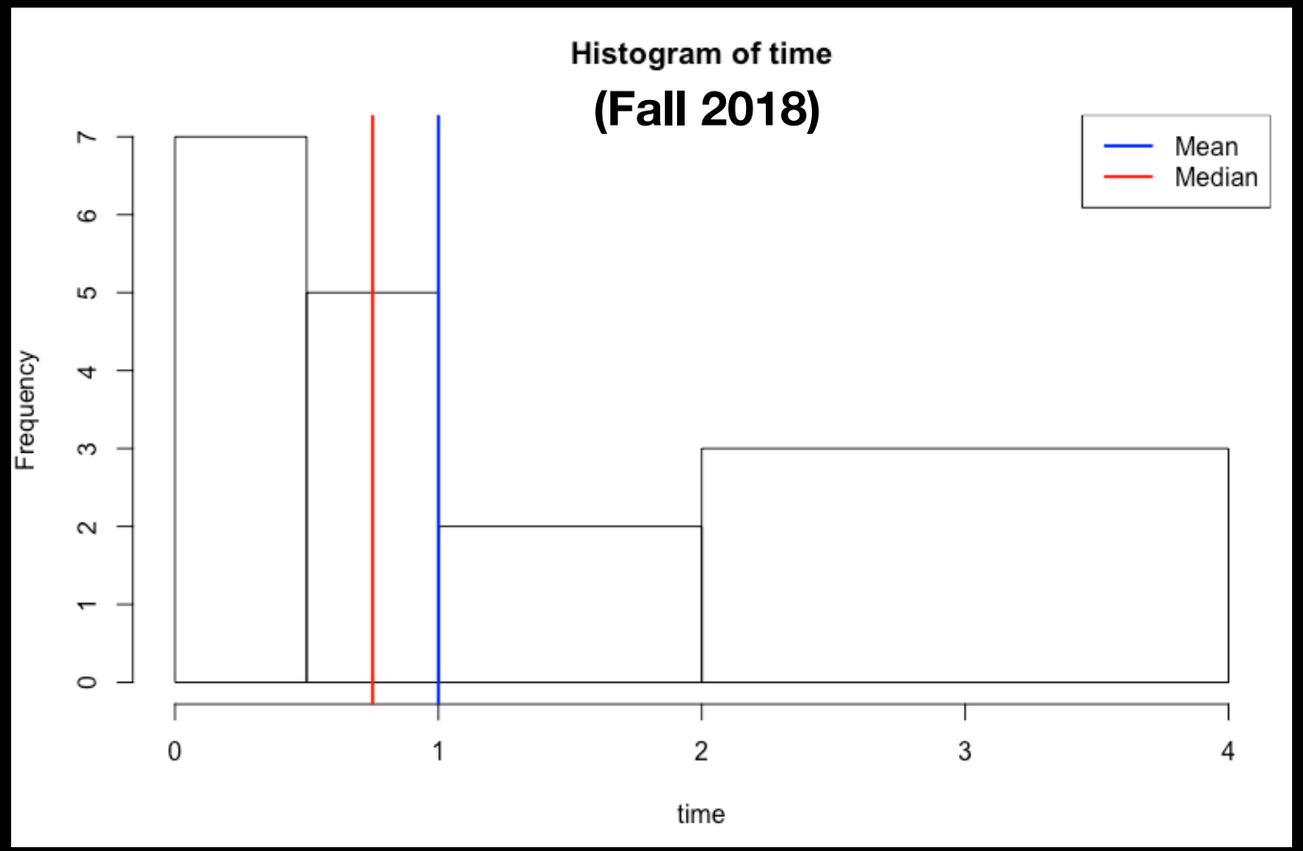
Median and quartile ranges are “robust” statistics - i.e. not as sensitive to variability

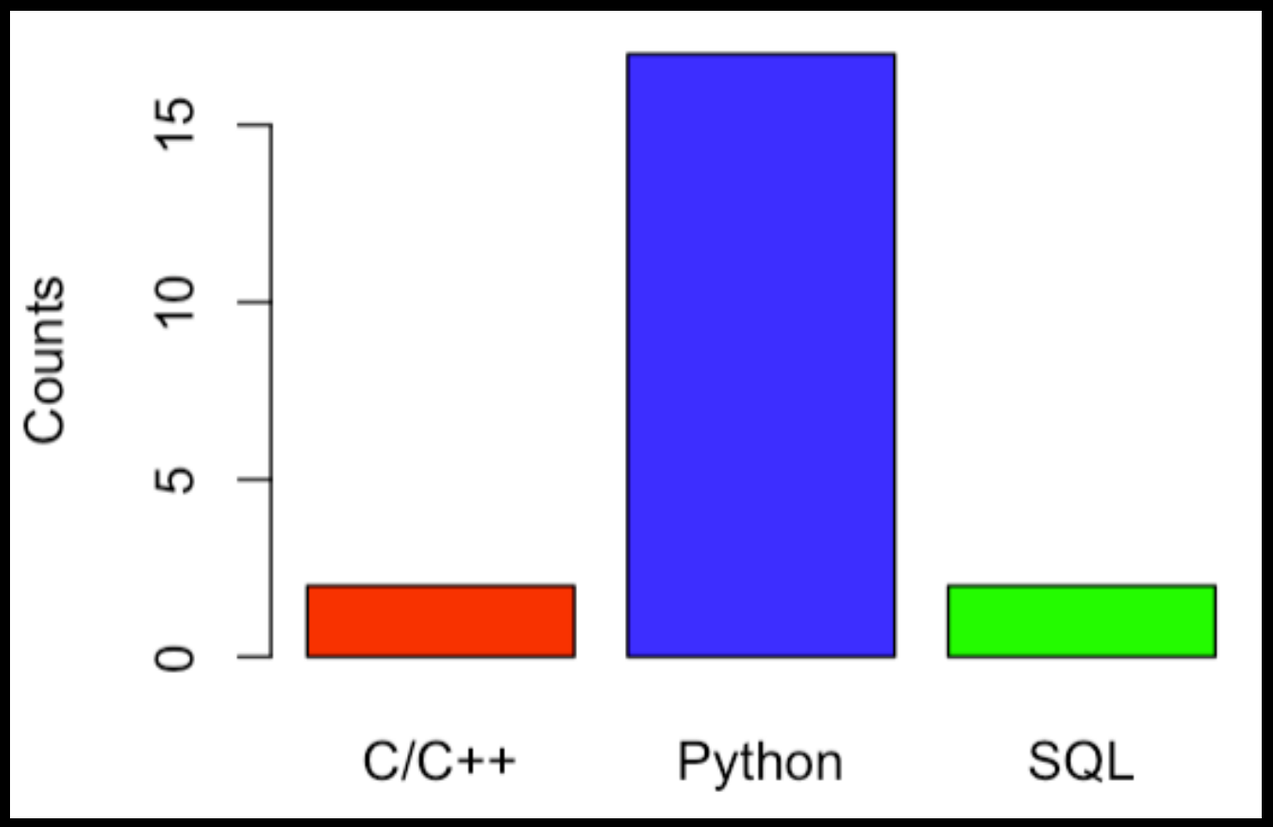



Histogram of time
(Fall 2018)
Mean
Median
Frequency
0
1
2
3
4
5
6
7
0
1
2
3
4
time


Histogram of time
(Fall 2018)
Mean
Median
Frequency
0
1
2
3
4
5
6
7
0
4

Left-Skewed
Right-Skewed


Histogram of time
(Fall 2018)
Mean
Median
Frequency
0
1
2
3
4
5
6
7
0
1
2
3
4
time


Counts
0
5
10
15
C/C++
Python
SQL

	Fam.Stats	language	time_yr.min	time_yr.max
1	4	Python	0.0	0.5
2	4	Python	2.0	4.0
3	3	Python	0.0	0.5
4	4	Matlab	2.0	4.0
5	2	Python	2.0	4.0
6	3	Python	1.0	2.0
7	3	Python	0.5	1.0
8	3	Python	0.5	1.0
9	3	C/C++	0.5	1.0
10	3	Java	0.0	0.5

Case

	Fam.Stats	language	time_yr.min	time_yr.max
1	4	Python	0.0	0.5
2	4	Python	2.0	4.0
3	3	Python	0.0	0.5
4	4	Matlab	2.0	4.0
5	2	Python	2.0	4.0
6	3	Python	1.0	2.0
7	3	Python	0.5	1.0
8	3	Python	0.5	1.0
9	3	C/C++	0.5	1.0
10	3	Java	0.0	0.5

Variables

Case

	Fam.Stats	language	time_yr.min	time_yr.max
1	4	Python	0.0	0.5
2	4	Python	2.0	4.0
3	3	Python	0.0	0.5
4	4	Matlab	2.0	4.0
5	2	Python	2.0	4.0
6	3	Python	1.0	2.0
7	3	Python	0.5	1.0
8	3	Python	0.5	1.0
9	3	C/C++	0.5	1.0
10	3	Java	0.0	0.5

Variables

Data Matrix

Case

	Fam.Stats	language	time_yr.min	time_yr.max
1	4	Python	0.0	0.5
2	4	Python	2.0	4.0
3	3	Python	0.0	0.5
4	4	Matlab	2.0	4.0
5	2	Python	2.0	4.0
6	3	Python	1.0	2.0
7	3	Python	0.5	1.0
8	3	Python	0.5	1.0
9	3	C/C++	0.5	1.0
10	3	Java	0.0	0.5

Variable Types

	Fam.Stats	language	time_yr.min	time_yr.max
1	4	Python	0.0	0.5
2	4	Python	2.0	4.0
3	3	Python	0.0	0.5
4	4	Matlab	2.0	4.0
5	2	Python	2.0	4.0
6	3	Python	1.0	2.0
7	3	Python	0.5	1.0
8	3	Python	0.5	1.0
9	3	C/C++	0.5	1.0
10	3	Java	0.0	0.5

Variable Types

Numerical
(+/-, means, etc)

	Fam.Stats	language	time_yr.min	time_yr.max
1	4	Python	0.0	0.5
2	4	Python	2.0	4.0
3	3	Python	0.0	0.5
4	4	Matlab	2.0	4.0
5	2	Python	2.0	4.0
6	3	Python	1.0	2.0
7	3	Python	0.5	1.0
8	3	Python	0.5	1.0
9	3	C/C++	0.5	1.0
10	3	Java	0.0	0.5

Variable Types

NOT Numerical

Numerical
(+/-, means, etc)

	Fam.Stats	language	time_yr.min	time_yr.max
1	4	Python	0.0	0.5
2	4	Python	2.0	4.0
3	3	Python	0.0	0.5
4	4	Matlab	2.0	4.0
5	2	Python	2.0	4.0
6	3	Python	1.0	2.0
7	3	Python	0.5	1.0
8	3	Python	0.5	1.0
9	3	C/C++	0.5	1.0
10	3	Java	0.0	0.5

Variable Types

NOT Numerical

Numerical
(+/-, means, etc)

	Fam.Stats	language	time_yr.min	time_yr.max
1	4	Python	0.0	0.5
2	4	Python	2.0	4.0
3	3	Python	0.0	0.5
4	4	Matlab	2.0	4.0
5	2	Python	2.0	4.0
6	3	Python	1.0	2.0
7	3	Python	0.5	1.0
8	3	Python	0.5	1.0
9	3	C/C++	0.5	1.0
10	3	Java	0.0	0.5

discrete numerical

vs.

continuous numerical

Variable Types

Categorical (levels)

Numerical (+/-, means, etc)

	Fam.Stats	language	time_yr.min	time_yr.max
1	4	Python	0.0	0.5
2	4	Python	2.0	4.0
3	3	Python	0.0	0.5
4	4	Matlab	2.0	4.0
5	2	Python	2.0	4.0
6	3	Python	1.0	2.0
7	3	Python	0.5	1.0
8	3	Python	0.5	1.0
9	3	C/C++	0.5	1.0
10	3	Java	0.0	0.5

Variable Types

Categorical (levels)

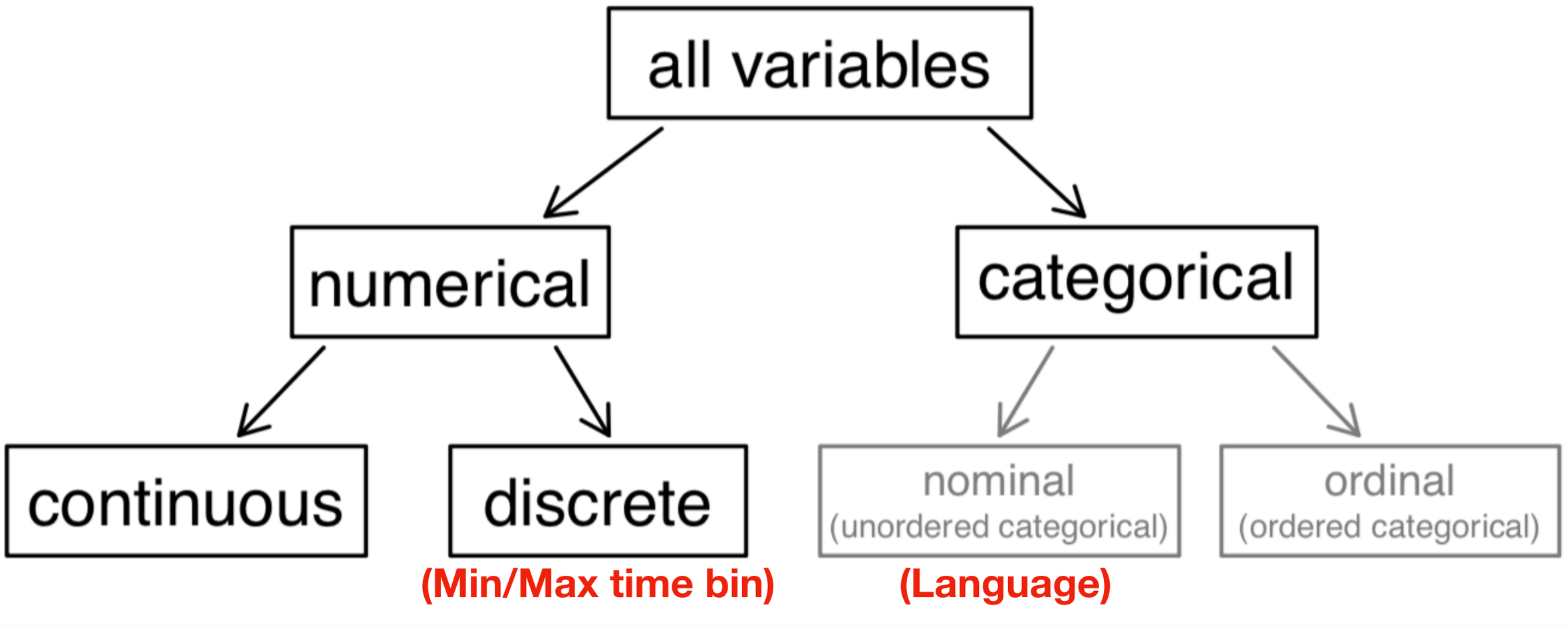
Numerical (+/-, means, etc)

	Fam.Stats	language	time_yr.min	time_yr.max
1	4	Python	0.0	0.5
2	4	Python	2.0	4.0
3	3	Python	0.0	0.5
4	4	Matlab	2.0	4.0
5	2	Python	2.0	4.0
6	3	Python	1.0	2.0
7	3	Python	0.5	1.0
8	3	Python	0.5	1.0
9	3	C/C++	0.5	1.0
10	3	Java	0.0	0.5

nominal categorical

vs.

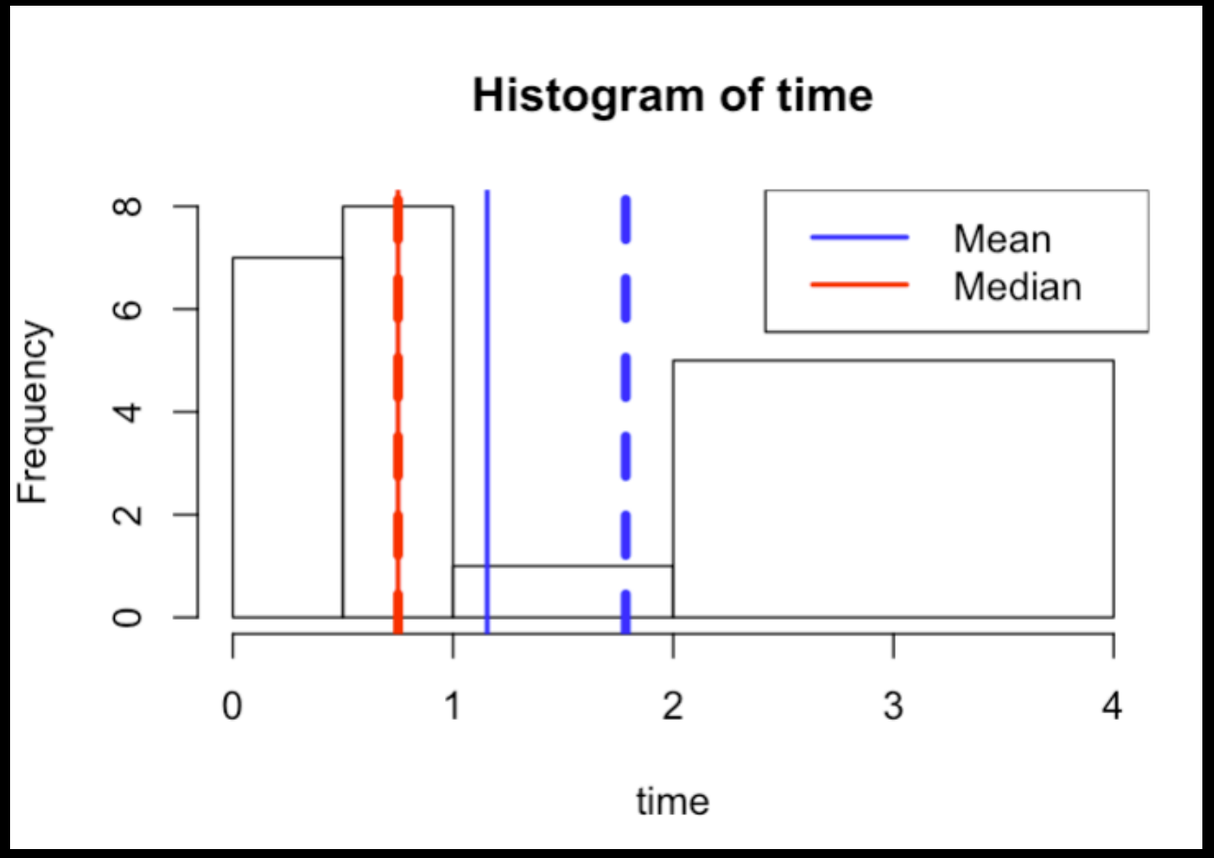
ordinal categorical

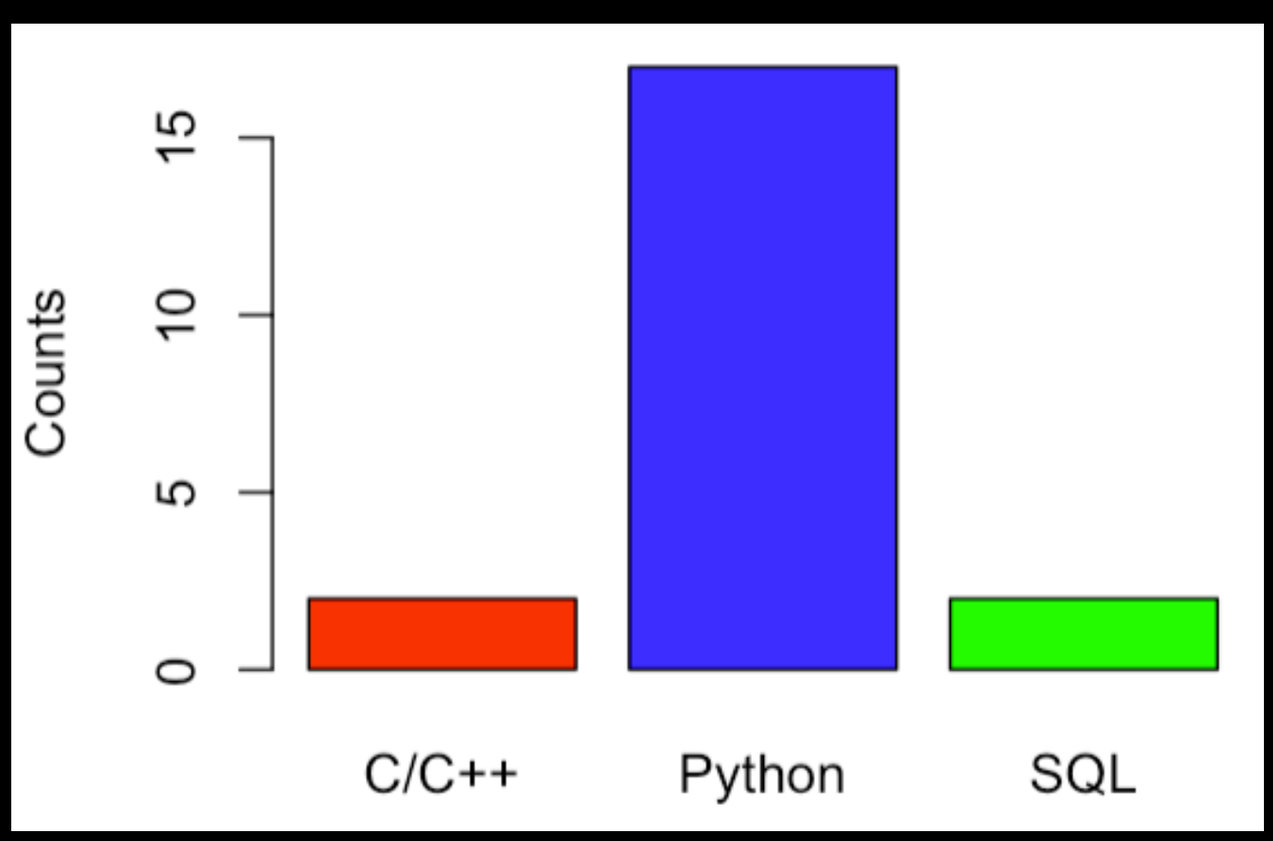

all variables
numerical
categorical
continuous
discrete
nominal
(unordered categorical)
ordinal
(ordered categorical)
(Min/Max time bin)
(Language)

Can we use this data to ask the following questions about the typical iSchool student?

1. **Can we say typical iSchool student codes in Python?**
2. **Can we say typical iSchool student has been programming for <1 year? >2 years?**

Why or why not?



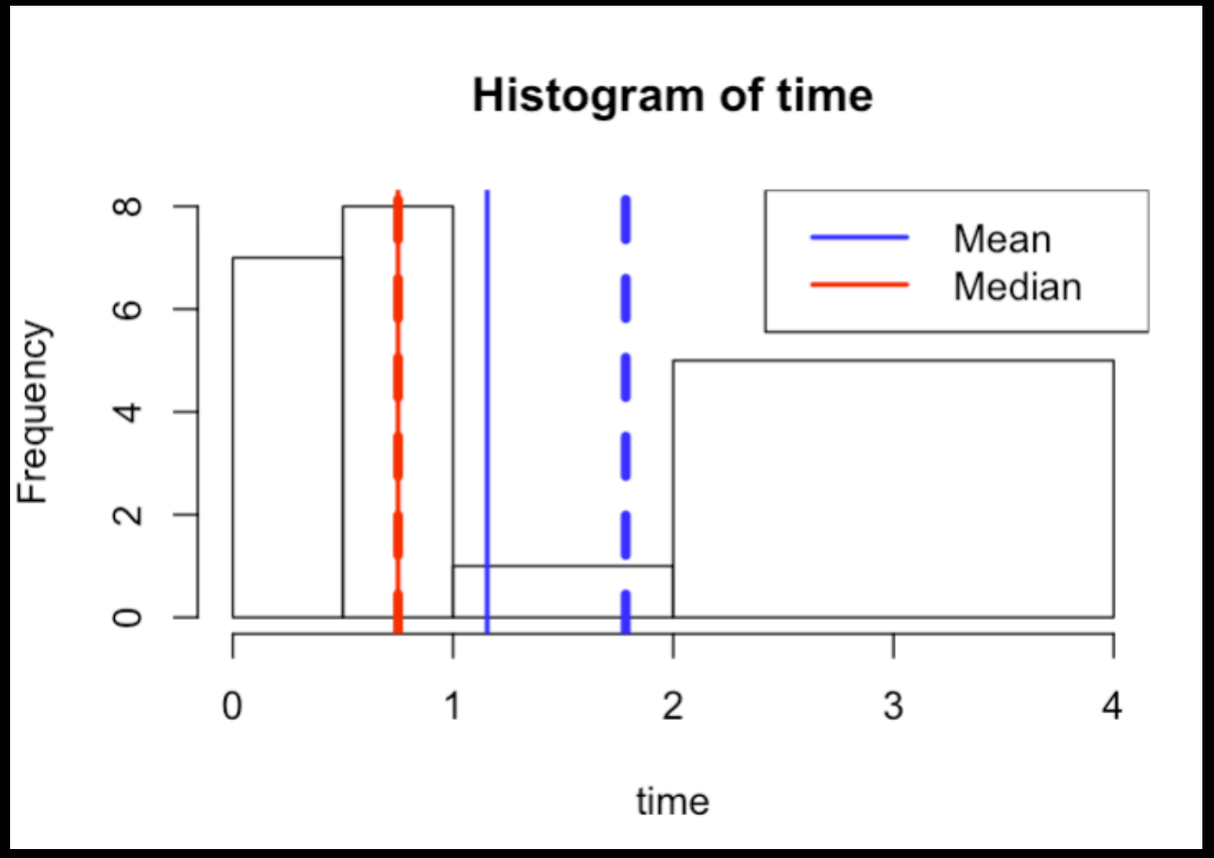


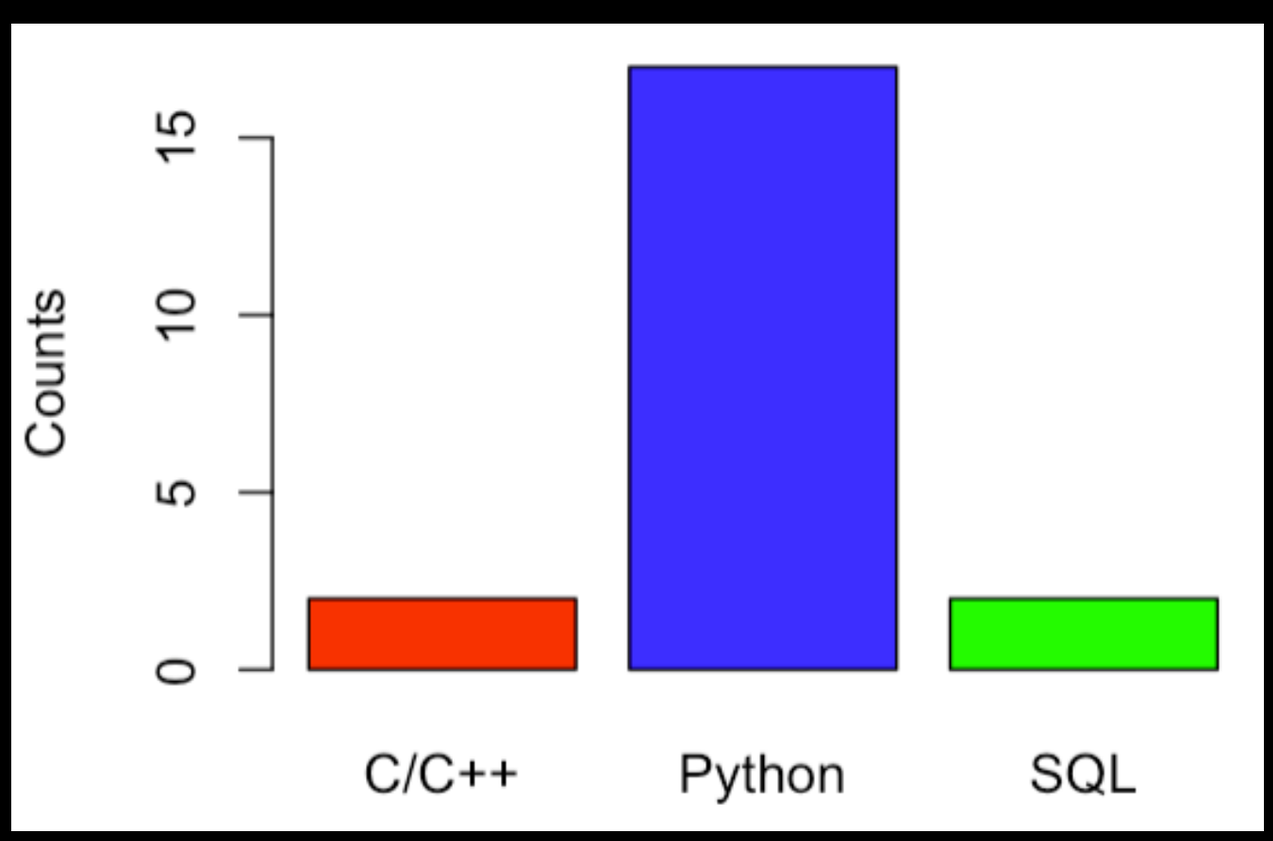
Can we use this data to ask the following questions about the typical iSchool student?

1. **Can we say typical iSchool student codes in python?**
2. **Can we say typical iSchool student has been programming for <1 year? >2 years?**

Why or why not?

Share: biases of sample?





Population of interest

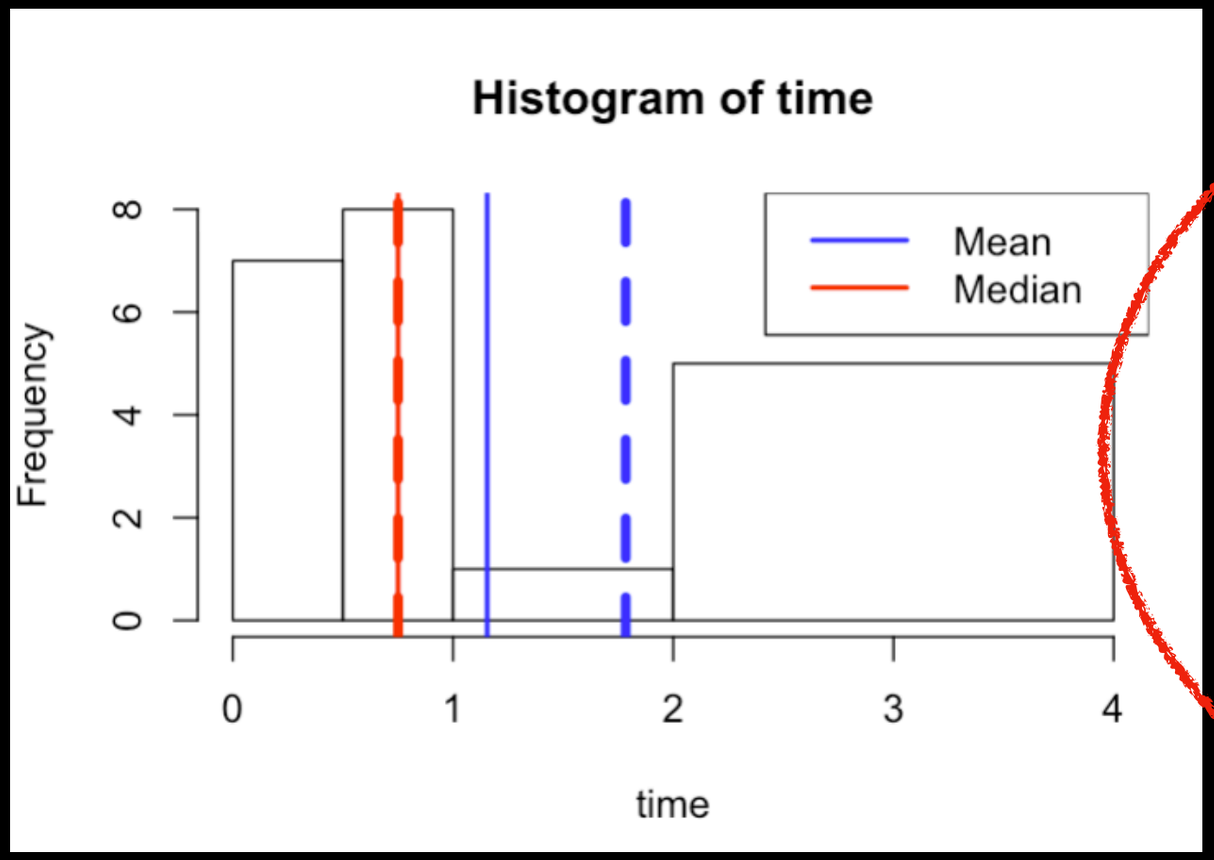
Can we use this data to ask the following questions about the typical iSchool student?

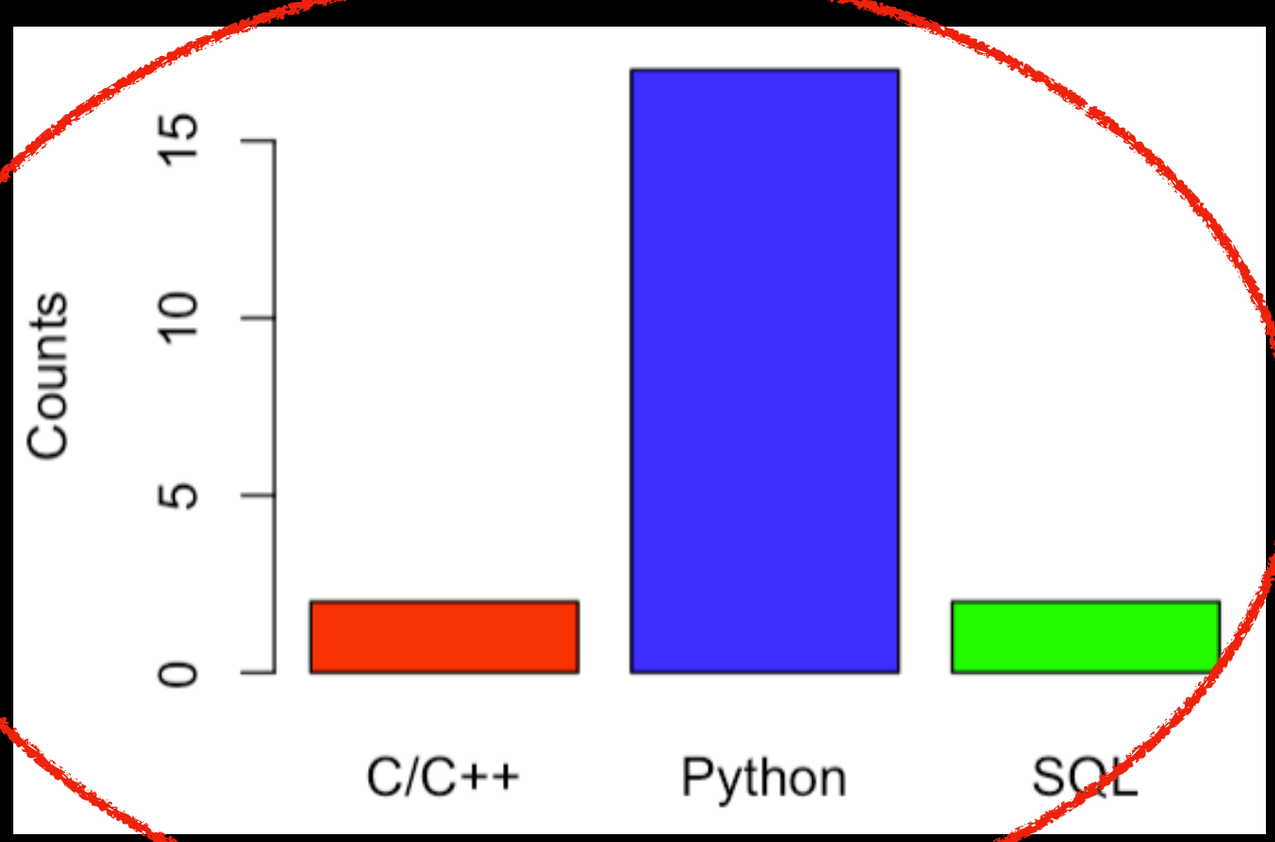
1. **Can we say typical iSchool student codes in python?**
2. **Can we say typical iSchool student has been programming for <1 year? >2 years?**

Why or why not?

Share: biases of sample?

Our survey was just a **sample**

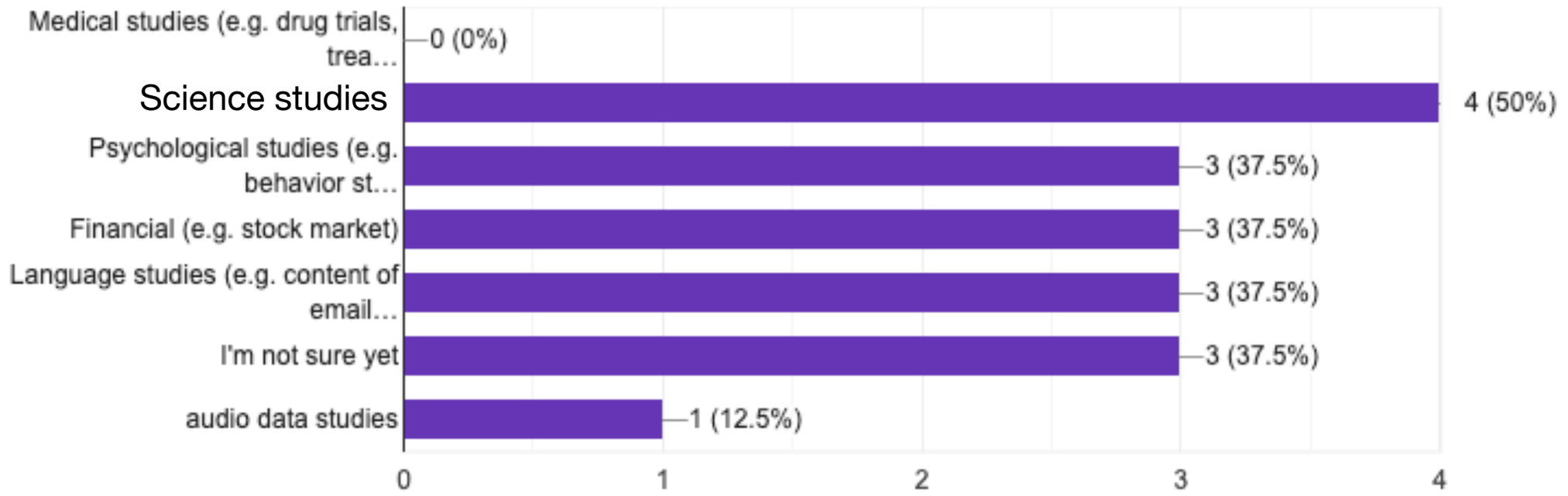




In fact a **convenience sample** (not a **random sample**)

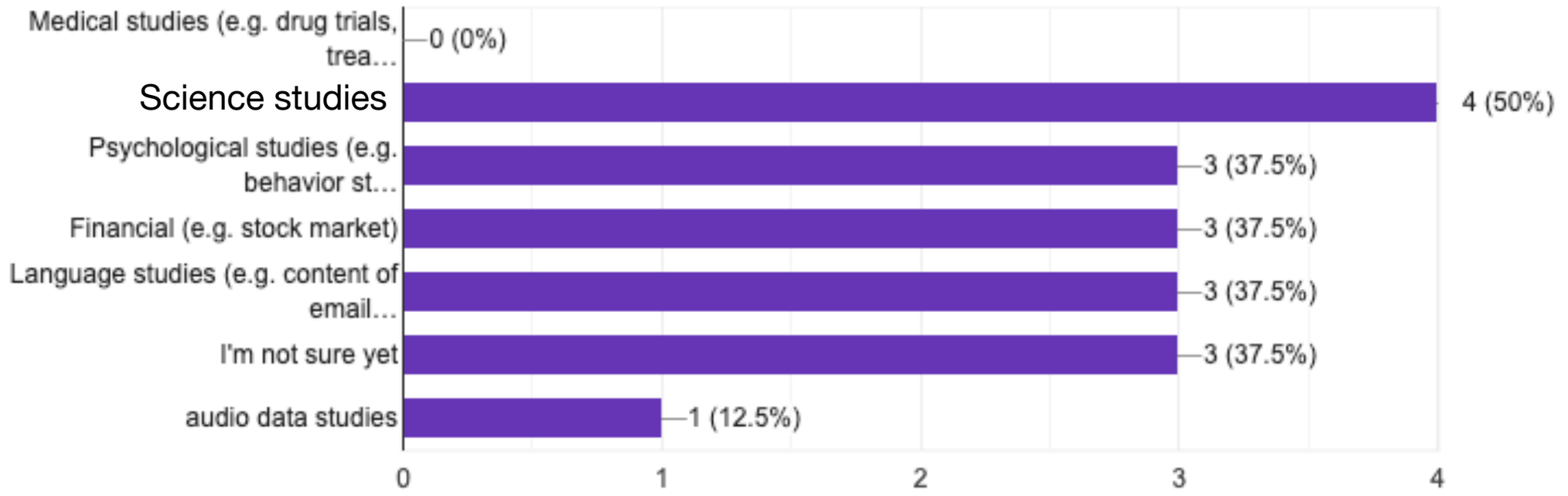
What kinds of datasets are you looking forward to visualizing? Check all that apply.

8 responses



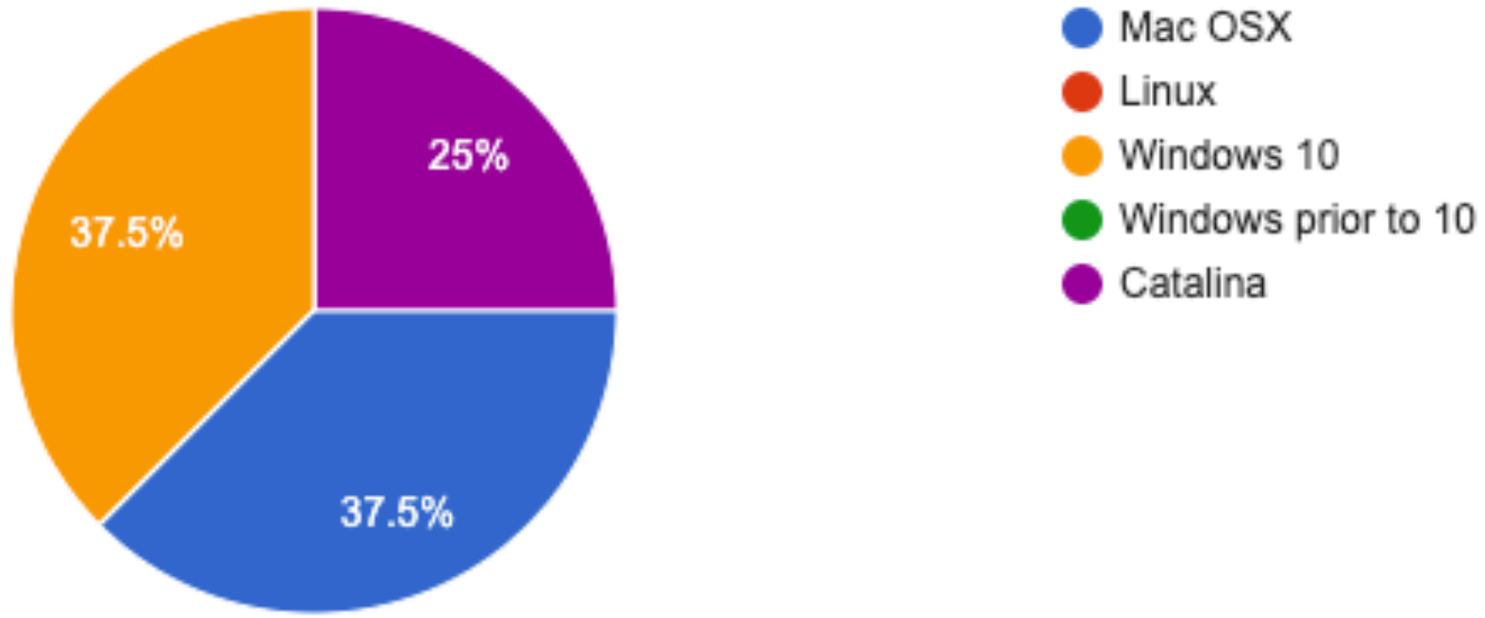
What kinds of datasets are you looking forward to visualizing? Check all th

8 responses



Which operation system do you use on your laptop?

8 responses



Real Data! (ooooo!)

Dataset #1:

http://data.un.org/

to

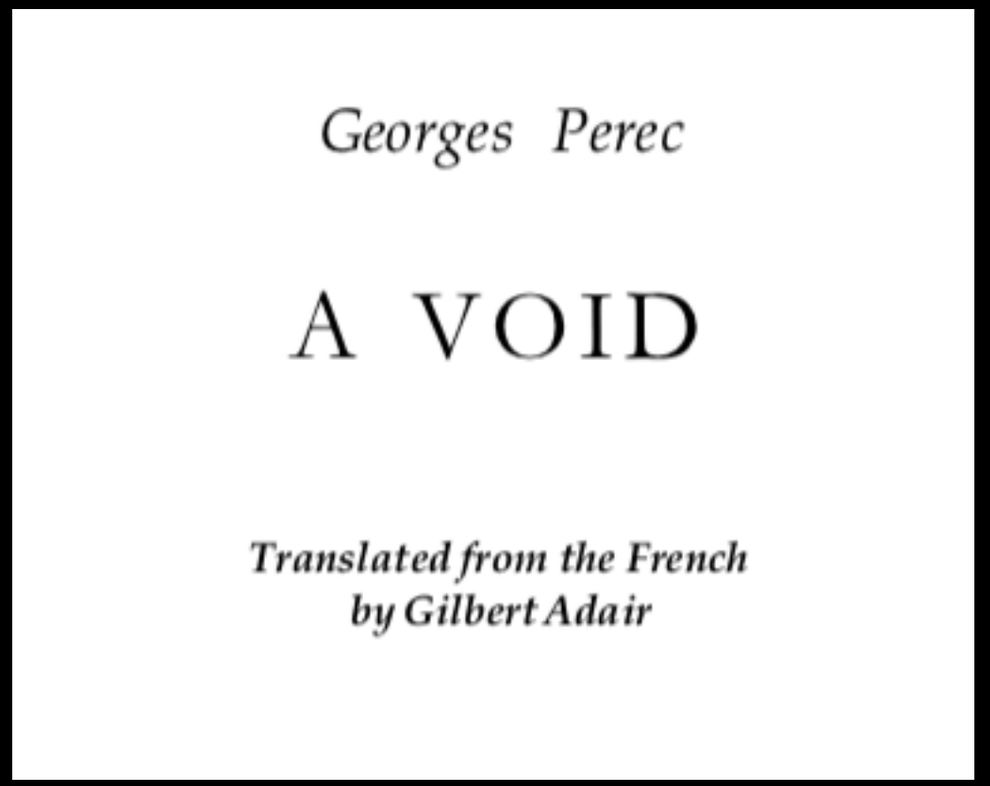
http://data.un.org/Explorer.aspx

to

Commodity trade stats =>
Fish, crustaceans, mollusks, etc

Dataset #2:

Excerpt of the
Introduction of “A Void”
by Georges Perec



Dataset #1

http://data.un.org/Data.aspx?d=ComTrade&f=_l1Code%3a4

or

https://tinyurl.com/y8vo7v8u

Download | Explore | Select columns | Select sort order | Link to this page

336777 records | Page 1 of 6736 |

Country or Area	Year	Commodity	Flow	Trade (USD)	Weight (kg)	Quantity Name	Quantity
Afghanistan	2010	Trout, fresh or chilled, whole	Import	8,600	9,000	Weight in kilograms	9,000
Albania	2016	Fish live, except trout, eel or carp	Import	2,202,944	39,896	Weight in kilograms	39,896
Albania	2016	Trout, fresh or chilled, whole	Export	1,973,381	266,283	Weight in kilograms	266,283
Albania	2016	Salmon fresh or chilled, whole	Import	387,938	96,083	Weight in kilograms	96,083
Albania	2016	Salmon fresh or chilled, whole	Export	39,162	5,974	Weight in kilograms	5,974
Albania	2016	Salmonidae, not trout or salmon,fresh or chilled whol	Import	70,993	11,065	Weight in kilograms	11,065
Albania	2016	Salmonidae, not trout or salmon,fresh or chilled whol	Export	49,287	10,808	Weight in kilograms	10,808
Albania	2016	Sardines,brisling,sprats, fresh or chilled, whole	Import	305,172	457,234	Weight in kilograms	457,234
Albania	2016	Fish nes, fresh or chilled, whole	Import	2,654,089	1,322,338	Weight in kilograms	1,322,338
Albania	2016	Fish nes, fresh or chilled, whole	Export	1,353,539	241,436	Weight in kilograms	241,436
Albania	2016	Fish livers and roes, fresh or chilled	Import	3,525	51	Weight in kilograms	51
Albania	2016	Salmon Atlantic or Danube, frozen, whole	Export	22,526	2,445	Weight in kilograms	2,445
Albania	2016	Fish nes, frozen, whole	Import	1,114,294	698,868	Weight in kilograms	698,868
Albania	2016	Fish nes, frozen, whole	Export	631,255	119,205	Weight in kilograms	119,205
Albania	2016	Fish fillets, frozen	Export	147,044	22,523	Weight in kilograms	22,523
Albania	2016	Fish meat & mince, except liver, roe & fillets, froze	Export	111,045	20,212	Weight in kilograms	20,212
Albania	2016	Fish fillets, dried, salted or in brine, not smoked	Export	7,529,293	558,797	Weight in kilograms	558,797
Albania	2016	Salmon, smoked, including fillets	Import	14,497	779	Weight in kilograms	779
Albania	2016	Anchovies, salted or in brine, not dried or smoked	Import	18,934,237	5,799,550	Weight in kilograms	5,799,550

Dataset #1

http://data.un.org/Data.aspx?d=ComTrade&f=_l1Code%3a4

or

https://tinyurl.com/y8vo7v8u

Download | Explore | Select columns | Select sort order | Link to this page

336777 records | Page 1 of 6736 |

Country or Area	Year	Commodity	Flow	Trade (USD)	Weight (kg)	Quantity Name	Quantity
Afghanistan	2010	Trout, fresh or chilled, whole	Import	8,600	9,000	Weight in kilograms	9,000
Albania	2016	Fish live, except trout, eel or carp	Import	2,202,944	39,896	Weight in kilograms	39,896
Albania	2016	Trout, fresh or chilled, whole	Export	1,973,381	266,283	Weight in kilograms	266,283
Albania	2016	Salmon fresh or chilled, whole	Import	387,938	96,083	Weight in kilograms	96,083
Albania	2016	Salmon fresh or chilled, whole	Export	39,162	5,974	Weight in kilograms	5,974
Albania	2016	Salmonidae, not trout or salmon,fresh or chilled whol	Import	70,993	11,065	Weight in kilograms	11,065
Albania	2016	Salmonidae, not trout or salmon,fresh or chilled whol	Export	49,287	10,808	Weight in kilograms	10,808
Albania	2016	Sardines,brisling,sprats, fresh or chilled, whole	Import	305,172	457,234	Weight in kilograms	457,234
Albania	2016	Fish nes, fresh or chilled, whole	Import	2,654,089	1,322,338	Weight in kilograms	1,322,338
Albania	2016	Fish nes, fresh or chilled, whole	Export	1,353,539	241,436	Weight in kilograms	241,436
Albania	2016	Fish livers and roes, fresh or chilled	Import	3,525	51	Weight in kilograms	51
Albania	2016	Salmon Atlantic or Danube, frozen, whole	Export	22,526	2,445	Weight in kilograms	2,445
Albania	2016	Fish nes, frozen, whole	Import	1,114,294	698,868	Weight in kilograms	698,868
Albania	2016	Fish nes, frozen, whole	Export	631,255	119,205	Weight in kilograms	119,205
Albania	2016	Fish fillets, frozen	Export	147,044	22,523	Weight in kilograms	22,523
Albania	2016	Fish meat & mince, except liver, roe & fillets, froze	Export	111,045	20,212	Weight in kilograms	20,212
Albania	2016	Fish fillets, dried, salted or in brine, not smoked	Export	7,529,293	558,797	Weight in kilograms	558,797
Albania	2016	Salmon, smoked, including fillets	Import	14,497	779	Weight in kilograms	779
Albania	2016	Anchovies, salted or in brine, not dried or smoked	Import	18,934,237	5,799,550	Weight in kilograms	5,799,550

Discuss:
What is the population?
What is a case?
What are each type of variable (numerical, categorical, etc)

Note: could be multiple answers!

Dataset #1

Download | Explore | Select columns | Select sort order | Link to this page

336777 records | Page 1 of 6736 |

Country or Area	Year	Commodity	Flow	Trade (USD)	Weight (kg)	Quantity Name	Quantity
Afghanistan	2010	Trout, fresh or chilled, whole	Import	8,600	9,000	Weight in kilograms	9,000
Albania	2016	Fish live, except trout, eel or carp	Import	2,202,944	39,896	Weight in kilograms	39,896
Albania	2016	Trout, fresh or chilled, whole	Export	1,973,381	266,283	Weight in kilograms	266,283
Albania	2016	Salmon fresh or chilled, whole	Import	387,938	96,083	Weight in kilograms	96,083
Albania	2016	Salmon fresh or chilled, whole	Export	39,162	5,974	Weight in kilograms	5,974
Albania	2016	Salmonidae, not trout or salmon,fresh or chilled whol	Import	70,993	11,065	Weight in kilograms	11,065
Albania	2016	Salmonidae, not trout or salmon,fresh or chilled whol	Export	49,287	10,808	Weight in kilograms	10,808
Albania	2016	Sardines,brisling,sprats, fresh or chilled, whole	Import	305,172	457,234	Weight in kilograms	457,234
Albania	2016	Fish nes, fresh or chilled, whole	Import	2,654,089	1,322,338	Weight in kilograms	1,322,338
Albania	2016	Fish nes, fresh or chilled, whole	Export	1,353,539	241,436	Weight in kilograms	241,436
Albania	2016	Fish livers and roes, fresh or chilled	Import	3,525	51	Weight in kilograms	51
Albania	2016	Salmon Atlantic or Danube, frozen, whole	Export	22,526	2,445	Weight in kilograms	2,445
Albania	2016	Fish nes, frozen, whole	Import	1,114,294	698,868	Weight in kilograms	698,868
Albania	2016	Fish nes, frozen, whole	Export	631,255	119,205	Weight in kilograms	119,205
Albania	2016	Fish fillets, frozen	Export	147,044	22,523	Weight in kilograms	22,523
Albania	2016	Fish meat & mince, except liver, roe & fillets, froze	Export	111,045	20,212	Weight in kilograms	20,212
Albania	2016	Fish fillets, dried, salted or in brine, not smoked	Export	7,529,293	558,797	Weight in kilograms	558,797
Albania	2016	Salmon, smoked, including fillets	Import	14,497	779	Weight in kilograms	779
Albania	2016	Anchovies, salted or in brine, not dried or smoked	Import	18,934,237	5,799,550	Weight in kilograms	5,799,550

Poll:
Type of variable “Country or Area”

1. Continuous Numerical
2. Discrete Numerical
3. Nominal (unordered) Categorical
4. Ordinal (ordered) Categorical

Dataset #1

Download | Explore | Select columns | Select sort order | Link to this page

336777 records | Page 1 of 6736 |

Country or Area	Year	Commodity	Flow	Trade (USD)	Weight (kg)	Quantity Name	Quantity
Afghanistan	2010	Trout, fresh or chilled, whole	Import	8,600	9,000	Weight in kilograms	9,000
Albania	2016	Fish live, except trout, eel or carp	Import	2,202,944	39,896	Weight in kilograms	39,896
Albania	2016	Trout, fresh or chilled, whole	Export	1,973,381	266,283	Weight in kilograms	266,283
Albania	2016	Salmon fresh or chilled, whole	Import	387,938	96,083	Weight in kilograms	96,083
Albania	2016	Salmon fresh or chilled, whole	Export	39,162	5,974	Weight in kilograms	5,974
Albania	2016	Salmonidae, not trout or salmon,fresh or chilled whol	Import	70,993	11,065	Weight in kilograms	11,065
Albania	2016	Salmonidae, not trout or salmon,fresh or chilled whol	Export	49,287	10,808	Weight in kilograms	10,808
Albania	2016	Sardines,brisling,sprats, fresh or chilled, whole	Import	305,172	457,234	Weight in kilograms	457,234
Albania	2016	Fish nes, fresh or chilled, whole	Import	2,654,089	1,322,338	Weight in kilograms	1,322,338
Albania	2016	Fish nes, fresh or chilled, whole	Export	1,353,539	241,436	Weight in kilograms	241,436
Albania	2016	Fish livers and roes, fresh or chilled	Import	3,525	51	Weight in kilograms	51
Albania	2016	Salmon Atlantic or Danube, frozen, whole	Export	22,526	2,445	Weight in kilograms	2,445
Albania	2016	Fish nes, frozen, whole	Import	1,114,294	698,868	Weight in kilograms	698,868
Albania	2016	Fish nes, frozen, whole	Export	631,255	119,205	Weight in kilograms	119,205
Albania	2016	Fish fillets, frozen	Export	147,044	22,523	Weight in kilograms	22,523
Albania	2016	Fish meat & mince, except liver, roe & fillets, froze	Export	111,045	20,212	Weight in kilograms	20,212
Albania	2016	Fish fillets, dried, salted or in brine, not smoked	Export	7,529,293	558,797	Weight in kilograms	558,797
Albania	2016	Salmon, smoked, including fillets	Import	14,497	779	Weight in kilograms	779
Albania	2016	Anchovies, salted or in brine, not dried or smoked	Import	18,934,237	5,799,550	Weight in kilograms	5,799,550

Poll:
Type of variable “Trade USD”

1. Continuous Numerical
2. Discrete Numerical
3. Nominal (unordered) Categorical
4. Ordinal (ordered) Categorical

Dataset #1

Download | Explore | Select columns | Select sort order | Link to this page

336777 records | Page 1 of 6736 |

Country or Area	Year	Commodity	Flow	Trade (USD)	Weight (kg)	Quantity Name	Quantity
Afghanistan	2010	Trout, fresh or chilled, whole	Import	8,600	9,000	Weight in kilograms	9,000
Albania	2016	Fish live, except trout, eel or carp	Import	2,202,944	39,896	Weight in kilograms	39,896
Albania	2016	Trout, fresh or chilled, whole	Export	1,973,381	266,283	Weight in kilograms	266,283
Albania	2016	Salmon fresh or chilled, whole	Import	387,938	96,083	Weight in kilograms	96,083
Albania	2016	Salmon fresh or chilled, whole	Export	39,162	5,974	Weight in kilograms	5,974
Albania	2016	Salmonidae, not trout or salmon,fresh or chilled whol	Import	70,993	11,065	Weight in kilograms	11,065
Albania	2016	Salmonidae, not trout or salmon,fresh or chilled whol	Export	49,287	10,808	Weight in kilograms	10,808
Albania	2016	Sardines,brisling,sprats, fresh or chilled, whole	Import	305,172	457,234	Weight in kilograms	457,234
Albania	2016	Fish nes, fresh or chilled, whole	Import	2,654,089	1,322,338	Weight in kilograms	1,322,338
Albania	2016	Fish nes, fresh or chilled, whole	Export	1,353,539	241,436	Weight in kilograms	241,436
Albania	2016	Fish livers and roes, fresh or chilled	Import	3,525	51	Weight in kilograms	51
Albania	2016	Salmon Atlantic or Danube, frozen, whole	Export	22,526	2,445	Weight in kilograms	2,445
Albania	2016	Fish nes, frozen, whole	Import	1,114,294	698,868	Weight in kilograms	698,868
Albania	2016	Fish nes, frozen, whole	Export	631,255	119,205	Weight in kilograms	119,205
Albania	2016	Fish fillets, frozen	Export	147,044	22,523	Weight in kilograms	22,523
Albania	2016	Fish meat & mince, except liver, roe & fillets, froze	Export	111,045	20,212	Weight in kilograms	20,212
Albania	2016	Fish fillets, dried, salted or in brine, not smoked	Export	7,529,293	558,797	Weight in kilograms	558,797
Albania	2016	Salmon, smoked, including fillets	Import	14,497	779	Weight in kilograms	779
Albania	2016	Anchovies, salted or in brine, not dried or smoked	Import	18,934,237	5,799,550	Weight in kilograms	5,799,550

Poll:
Type of variable “Year”

1. Continuous Numerical
2. Discrete Numerical
3. Nominal (unordered) Categorical
4. Ordinal (ordered) Categorical

Dataset #1

Download | Explore | Select columns | Select sort order | Link to this page

336777 records | Page 1 of 6736 |

Country or Area	Year	Commodity	Flow	Trade (USD)	Weight (kg)	Quantity Name	Quantity
Afghanistan	2010	Trout, fresh or chilled, whole	Import	8,600	9,000	Weight in kilograms	9,000
Albania	2016	Fish live, except trout, eel or carp	Import	2,202,944	39,896	Weight in kilograms	39,896
Albania	2016	Trout, fresh or chilled, whole	Export	1,973,381	266,283	Weight in kilograms	266,283
Albania	2016	Salmon fresh or chilled, whole	Import	387,938	96,083	Weight in kilograms	96,083
Albania	2016	Salmon fresh or chilled, whole	Export	39,162	5,974	Weight in kilograms	5,974
Albania	2016	Salmonidae, not trout or salmon,fresh or chilled whol	Import	70,993	11,065	Weight in kilograms	11,065
Albania	2016	Salmonidae, not trout or salmon,fresh or chilled whol	Export	49,287	10,808	Weight in kilograms	10,808
Albania	2016	Sardines,brisling,sprats, fresh or chilled, whole	Import	305,172	457,234	Weight in kilograms	457,234
Albania	2016	Fish nes, fresh or chilled, whole	Import	2,654,089	1,322,338	Weight in kilograms	1,322,338
Albania	2016	Fish nes, fresh or chilled, whole	Export	1,353,539	241,436	Weight in kilograms	241,436
Albania	2016	Fish livers and roes, fresh or chilled	Import	3,525	51	Weight in kilograms	51
Albania	2016	Salmon Atlantic or Danube, frozen, whole	Export	22,526	2,445	Weight in kilograms	2,445
Albania	2016	Fish nes, frozen, whole	Import	1,114,294	698,868	Weight in kilograms	698,868
Albania	2016	Fish nes, frozen, whole	Export	631,255	119,205	Weight in kilograms	119,205
Albania	2016	Fish fillets, frozen	Export	147,044	22,523	Weight in kilograms	22,523
Albania	2016	Fish meat & mince, except liver, roe & fillets, froze	Export	111,045	20,212	Weight in kilograms	20,212
Albania	2016	Fish fillets, dried, salted or in brine, not smoked	Export	7,529,293	558,797	Weight in kilograms	558,797
Albania	2016	Salmon, smoked, including fillets	Import	14,497	779	Weight in kilograms	779
Albania	2016	Anchovies, salted or in brine, not dried or smoked	Import	18,934,237	5,799,550	Weight in kilograms	5,799,550

What are some questions you might want to use your new R-skills (plotting, calculations) and summary statistics knowledge (mean, median, standard deviation, quartiles) to probe with this dataset?

Dataset #2

INTRODUCTION

In which, as you will soon find out, Damnation has its origin

Today, by radio, and also on giant hoardings, a rabbi, an admiral notorious for his links to Masonry, a trio of cardinals, a trio, too, of insignificant politicians (bought and paid for by a rich and corrupt Anglo-Canadian banking corporation), inform us all of how our country now risks dying of starvation. A rumour, that's my initial thought as I switch off my radio, a rumour or possibly a hoax. Propaganda, I murmur anxiously — as though, just by saying so, I might allay my doubts — typical politicians' propaganda. But public opinion gradually absorbs it as a fact. Individuals start strutting around with stout clubs. "Food, glorious food!" is a common cry (occasionally sung to Bart's music), with ordinary hard-working folk harassing officials, both local and national, and cursing capitalists and captains of industry. Cops shrink from going out on night shift. In Macon a mob storms a municipal building. In Rocadamour ruffians rob a hangar full of foodstuffs, pillaging tons of tuna fish, milk and cocoa, as also a vast quantity of corn - all of it, alas, totally unfit for human consumption. Without fuss or ado, and naturally without any sort of trial, an indignant crowd hangs 26 solicitors on a hastily built scaffold in front of Nancy's law courts (this Nancy is a town, not a woman) and ransacks a local journal, a disgusting right-wing rag that is siding against it. Up and down this land of ours looting has brought docks, shops and farms to a virtual standstill.

Discuss:
What is the population?
What is a case?
What are each type of variable (numerical, categorical, etc)

Note: could be multiple answers or somewhat nebulous answers!

Dataset #2

INTRODUCTION

In which, as you will soon find out, Damnation has its origin

Today, by radio, and also on giant hoardings, a rabbi, an admiral notorious for his links to Masonry, a trio of cardinals, a trio, too, of insignificant politicians (bought and paid for by a rich and corrupt Anglo-Canadian banking corporation), inform us all of how our country now risks dying of starvation. A rumour, that's my initial thought as I switch off my radio, a rumour or possibly a hoax. Propaganda, I murmur anxiously — as though, just by saying so, I might allay my doubts — typical politicians' propaganda. But public opinion gradually absorbs it as a fact. Individuals start strutting around with stout clubs. "Food, glorious food!" is a common cry (occasionally sung to Bart's music), with ordinary hard-working folk harassing officials, both local and national, and cursing capitalists and captains of industry. Cops shrink from going out on night shift. In Macon a mob storms a municipal building. In Rocadamour ruffians rob a hangar full of foodstuffs, pillaging tons of tuna fish, milk and cocoa, as also a vast quantity of corn - all of it, alas, totally unfit for human consumption. Without fuss or ado, and naturally without any sort of trial, an indignant crowd hangs 26 solicitors on a hastily built scaffold in front of Nancy's law courts (this Nancy is a town, not a woman) and ransacks a local journal, a disgusting right-wing rag that is siding against it. Up and down this land of ours looting has brought docks, shops and farms to a virtual standstill.

What are some questions you might want to use your new R-skills (plotting, calculations) and summary statistics knowledge (mean, median, standard deviation, quartiles) to probe with this dataset?

Real Data! (ooooo!)

We'll go through this one

Dataset #1:

http://data.un.org/

to

http://data.un.org/Explorer.aspx

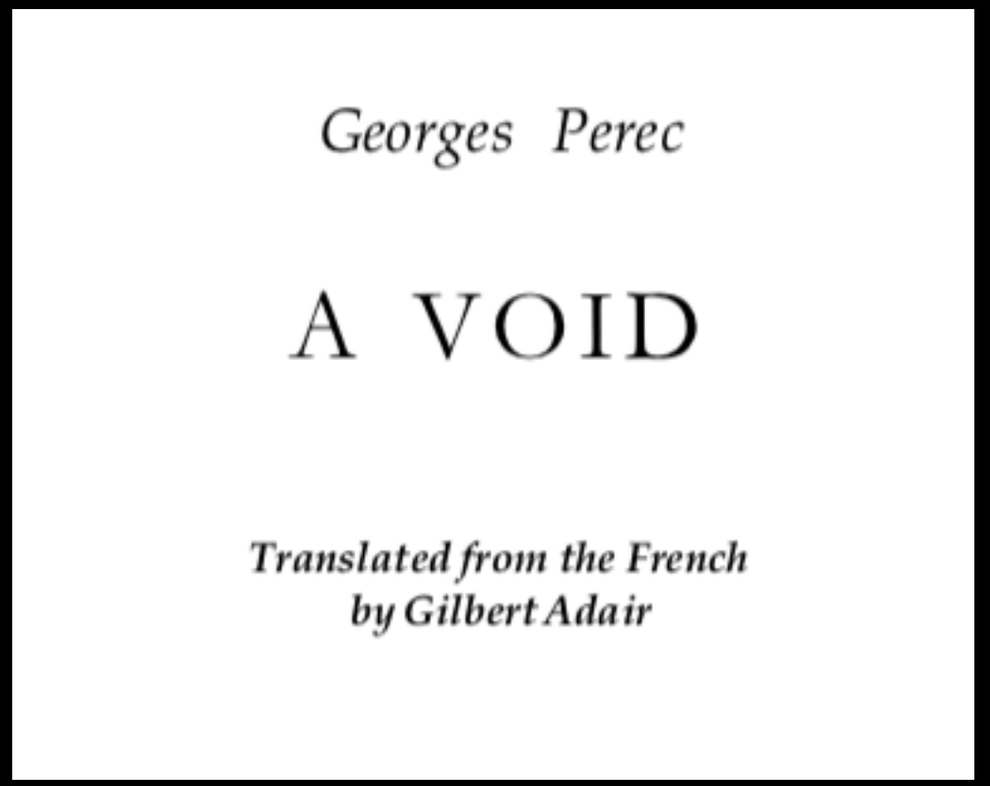
to

Commodity trade stats =>
Fish, crustaceans, mollusks, etc

We may or may not get to this one

Dataset #2:

Excerpt of the
Introduction of "A Void"
by Georges Perec



Real Data! (ooooo!)

Steps:

1. **Make sure everybody can download and open the data.**

2. **Make sure everybody can run the example code and look at the plots.**

3. **Come up with 1 (or more) questions/ideas/features you want to explore more with your dataset using R, plotting, summary statistics.**

4. **Write down the idea in the chat window so we can discuss as a group.**

5. **I'll go through some basic statistics things with this dataset - feel free to follow along or explore your idea. If exploring your own idea, please put in the chat window what you are doing and any questions you have.**

Real Data! (ooooo!)

Steps:

1. **Make sure everybody can download and open the data.**

2. **Make sure everybody can run the example code and look at the plots.**

3. **Come up with 1 (or more) questions/ideas/features you want to explore more with your dataset using R, plotting, summary statistics.**

4. **Write down the idea in the chat window so we can discuss as a group.**

5. **I'll go through some basic statistics things with this dataset - feel free to follow along or explore your idea. If exploring your own idea, please put in the chat window what you are doing and any questions you have.**

Confusion at this stage is TOTALLY normal!

Real Data! (ooooo!)

To R!

Steps:

1. **Make sure everybody can download and open the data.**

2. **Make sure everybody can run the example code and look at the plots.**

3. **Come up with 1 (or more) questions/ideas/features you want to explore more with your dataset using R, plotting, summary statistics.**

4. **Write down the idea in the chat window so we can discuss as a group.**

5. **I'll go through some basic statistics things with this dataset - feel free to follow along or explore your idea. If exploring your own idea, please put in the chat window what you are doing and any questions you have.**

Confusion at this stage is TOTALLY normal!