

# Exploring the Experiment's Design

In this set of exercises, we'll get our first look at the experiment we'll be analyzing in this course; curated data from the [Steinmetz et al, 2019](#) paper.

Today's data is focused on three CSV files, each containing sessions from a different stretch of data collection. They contain trial-level data from the experiment:

- steinmetz\_winter2016.csv
- steinmetz\_summer2017.csv
- steinmetz\_winter2017.csv

## Table of Contents

Loading and viewing data.....	1
Experiment Description: Calculating Statistics on Continuous Data.....	3
Experiment Description: Calculating Statistics of Across Categorical Data.....	4
Visualizing the Experimental Design using Plots.....	7

## Loading and viewing data

Let's load in some csv data with MATLAB and see what we have!

Code	Description
<code>readtable('my_datafile.csv')</code>	read a csv file
<code>data(row_start:row:end, :)</code>	select rows from row_start to row_end from data
<code>height(my_table)</code>	count the number of rows in my_table
<code>width(my_table)</code>	count the number of columns in my_table
<code>my_table.Properties.VariableNames</code>	get the column names of my_table

## Download datafiles

```
%cwd = fileparts(matlab.desktop.editor.getActiveFilename)
%url= "https://uni-bonn.sciebo.de/s/9FxelLhARmHpW85";
%mkdir(strcat(cwd, '/data'));
%webwrite(strcat(cwd, '/data/testfile.csv'), strcat(url, '/download') );
```

## Example:

Load in the winter 2016 dataset and view the first 5 rows of the data

```
data = readtable('data/steinmetz_winter2016.csv');
data(1:5, :)
```

```
ans = 5x15 table
```

...

	trial	active_trials	contrast_left	contrast_right	stim_onset
1	1	'True'	100	0	0.5000
2	2	'True'	0	50	0.5000
3	3	'True'	100	50	0.5000
4	4	'True'	0	0	0.5000
5	5	'True'	50	100	0.5000

The variable data is a MATLAB table, we can see that in the Workspace panel on the right.

## Exercises

Load in the winter 2016 dataset and view the first 10 rows of the data

```
data = readtable('data/steinmetz_winter2016_.csv');  
data(1:10,:)
```

```
ans = 10x16 table
```

...

	Var1	trial	active_trials	contrast_left	contrast_right
1	0	1	'True'	100	0
2	1	2	'True'	0	50
3	2	3	'True'	100	50
4	3	4	'True'	0	0
5	4	5	'True'	50	100
6	5	6	'True'	0	0
7	6	7	'True'	0	0
8	7	8	'True'	0	0
9	8	9	'True'	0	0
10	9	10	'True'	100	50

How many rows are in the summer 2016 dataset?

```
height(data)
```

```
ans = 2747
```

How many columns are in the summer 2016 dataset?

```
width(data)
```

```
ans = 15
```

What are the column names in the summer 2016 dataset?

```
data.Properties.VariableNames
```

```
ans = 1x15 cell  
'feedback_type' 'reaction_time' 'reaction_type' 'mouse' 'session_date' 'session_
```

Columns 10:15

## Experiment Description: Calculating Statistics on Continuous Data

Code	Description
my_table.column_A	access column_A of my_table
min(my_data)	find the minimum value of my_data
max(my_data)	find the maximum value of my_data
mean(my_data)	find the mean value of my_data
median(my_data)	find the media value of my_data
std(my_data)	find the standard deviation of my_data

### Example

Find the minimum response time in the table data

```
min(data.response_time)
```

```
ans = 0.5682
```

### Exercises

Find the maximum response time in the table data

```
max(data.response_time)
```

```
ans = 2.7136
```

Find the maximum gocue time in the table data

```
max(data.gocue_time)
```

```
ans = 1.1932
```

Find the minimum gocue time in the table data

```
min(data.gocue_time)
```

```
ans = 0.4804
```

Find the mean response time in the table data

```
mean(data.response_time)
```

```
ans = 1.5735
```

Find the median response\_time in the table data

```
median(data.response_time)
```

```
ans = 1.3541
```

Find the standard deviation of response time

```
std(data.response_time)
```

```
ans = 0.6402
```

Find the standard deviation of gocue time

```
std(data.gocue_time)
```

```
ans = 0.2011
```

## Experiment Description: Calculating Statistics of Across Categorical Data

In data science we often want to perform analysis on distinct separate categories, for example, analysing trials that occurred on different days. We can group the data and analyse each group separately, for example finding the mean response time for each mouse.

<u>Code</u>	<u>Description</u>
groupsummary(data, "column_A")	Group data according to column_A and count occurrences
groupsummary(data, "column_A", "mean", "column_B")	Group data according to column_A and calculate the mean of column_B

### Example Exercise

How many trials occurred for each session date?

```
groupsummary(data, "session_date")
```

```
ans = 9x2 table
```

	session_date	GroupCount
1	14-Dec-2016	214
2	17-Dec-2016	251
3	18-Dec-2016	228
4	07-Jan-2017	444
5	08-Jan-2017	412
6	09-Jan-2017	365
7	10-Jan-2017	253
8	11-Jan-2017	142
9	12-Jan-2017	128

How many trials did each mouse participate in?

```
groupsummary(data, "mouse")
```

```
ans = 3x2 table
```

	mouse	GroupCount
1	'Cori'	693
2	'Muller'	782
3	'Radnitz'	962

What was the mean response time for each mouse?

```
groupsummary(data, "mouse", "mean", "response_time")
```

```
ans = 3x3 table
```

	mouse	GroupCount	mean_response_time
1	'Cori'	693	1.5909
2	'Muller'	782	1.5758
3	'Radnitz'	962	1.5591

What was the minimum response time for each mouse?

```
groupsummary(data, "mouse", "min", "response_time")
```

```
ans = 3x3 table
```

	mouse	GroupCount	min_response_time
1	'Cori'	693	0.5682

	mouse	GroupCount	min_response_time
2	'Muller'	782	0.5854
3	'Radnitz'	962	0.5838

What was the most common (ie. mode) reaction\_type for each mouse?

```
groupsummary(data, "mouse", "mode", "reaction_type")
```

ans = 3x3 table

	mouse	GroupCount	mode_reaction_type
1	'Cori'	693	1
2	'Muller'	782	1
3	'Radnitz'	962	-1

What was the range of response times for each session date?

```
groupsummary(data, "session_date", "range", "response_time")
```

ans = 9x3 table

	session_date	GroupCount	range_response_time
1	14-Dec-2016	214	2.1054
2	17-Dec-2016	251	2.1227
3	18-Dec-2016	228	2.1076
4	07-Jan-2017	444	2.1245
5	08-Jan-2017	412	2.1158
6	09-Jan-2017	365	2.0917
7	10-Jan-2017	253	2.0796
8	11-Jan-2017	142	2.0306
9	12-Jan-2017	128	1.9787

Rerun the last exercise replacing "range" with "all"

```
groupsummary(data, "session_date", "all", "response_time")
```

ans = 9x14 table

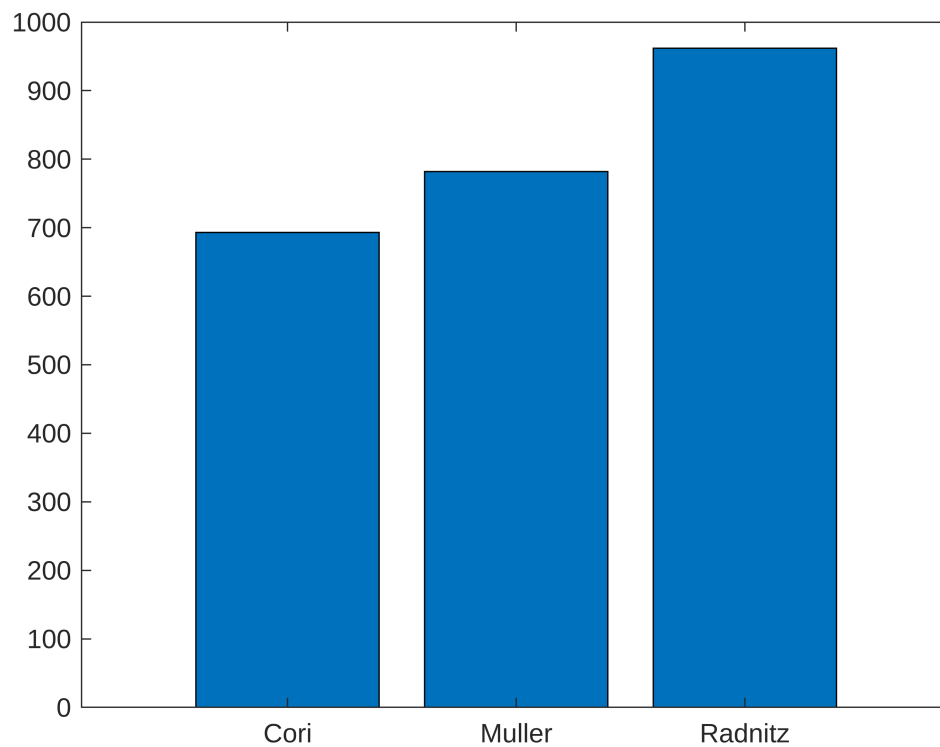
	session_date	GroupCount	mean_response_time	sum_response_time
1	14-Dec-2016	214	1.5891	340.0590
2	17-Dec-2016	251	1.6215	407.0063
3	18-Dec-2016	228	1.5588	355.4048

...

	session_date	GroupCount	mean_response_time	sum_response_time
4	07-Jan-2017	444	1.4704	652.8560
5	08-Jan-2017	412	1.5457	636.8091
6	09-Jan-2017	365	1.5928	581.3877
7	10-Jan-2017	253	1.5594	394.5252
8	11-Jan-2017	142	1.7116	243.0486
9	12-Jan-2017	128	1.7461	223.5022

## Visualizing the Experimental Design using Plots

```
trials_by_mouse = groupsummary(data, "mouse");
bar(trials_by_mouse.mouse, trials_by_mouse.GroupCount)
```



```
unique(data.session_date) % Note Matlab autoconverted session_dates to
datetime!
```

```
ans = 9x1 datetime
14-Dec-2016
17-Dec-2016
```

18-Dec-2016  
07-Jan-2017  
08-Jan-2017  
09-Jan-2017  
10-Jan-2017  
11-Jan-2017  
12-Jan-2017

```
trials_by_date = groupcounts(data, "session_date")
```

```
trials_by_date = 9x3 table
```

	session_date	GroupCount	Percent
1	14-Dec-2016	214	8.7813
2	17-Dec-2016	251	10.2995
3	18-Dec-2016	228	9.3558
4	07-Jan-2017	444	18.2191
5	08-Jan-2017	412	16.9060
6	09-Jan-2017	365	14.9774
7	10-Jan-2017	253	10.3816
8	11-Jan-2017	142	5.8268
9	12-Jan-2017	128	5.2524

```
histogram('Categories',  
unique(trials_by_date.day_session_date ), 'BinCounts',  
trials_by_date.GroupCount)
```

Error using .  
Unrecognized table variable name 'day\_session\_date'.

```
ylabel('num. trials')
```

```
bar( unique(trials_by_date.day_session_date ), trials_by_date.GroupCount)
```

Error using .  
Unrecognized table variable name 'day\_session\_date'.



