# ENGR 112: Engineering Decisions

Week # 6: Data Presentation and Decisions - Lab

DEPARTMENT OF ENGINEERING
JAMES MADISON UNIVERSITY

Spring 2022

Jason Forsyth, Ph.D.
Kyle Gipson, Ph.D.
Shraddha Joshi, Ph.D.
Steven Woodruff, ABD

# Outline

Examine several datasets: writing, heart EKG, and walking

All data pre-packaged into Excel files; download from Canvas in the Modules section.

Will walk through basic plots and analysis. Independent work on large dataset due at the end of the week.

# Recall: Engineers Collect A Lot of Data

- Information is collected to:
  - Convey operational information
  - Assess functionality
  - Examine relationships and trends
  - To select between alternatives
  - To determine effectiveness

- Overall, there is a purpose to generating/collecting information. In each project, ask yourself:
  - What data should I collect?
  - Is that information appropriate for the task?
  - Once collected, how might I manipulate it to answer questions and derive insights?

First Data Set. Read the description. What kind of data should we expect to see? (images, points, video…?) What units will that information have?

# Complex Upper-Limb Movements

## Introduction

The Complex Upper-Limb Movements database contains hand trajectory data collected from ten subjects as they performed various upper-limb motor tasks. The data was used in the above-referenced manuscript to identify the motor primitives contributing to the observed motor patterns.

Ten healthy subjects (7 males; 26.4±4.52 years of age; 9 right-handed) with no known neurological or orthopedic conditions affecting the control of motion were recruited to participate in the study. Written informed consent was obtained from all subjects. The protocol was approved by the Institutional Review Board of Spaulding Rehabilitation Hospital. The dataset was collected in the Motion Analysis Lab (http://srh-mal.net/) at Spaulding Rehabilitation Hospital.

The csv data files contain four columns. The data in the first column is the time axis in seconds. The data samples in the other three columns are the x-, y-, and z-coordinates of the reflective marker utilized in each experiment to represent the hand trajectory of movement. The data is reported in meters. Different files contain data pertaining to different subjects for different motor tasks. Different filenames are used for different tasks as follows:

https://physionet.org/content/culm/1.0.0/

# Critical Questions for the Dataset

- What type of information was collected?

- What are the units for each sample?

- What should the data look like?

# Critical Questions for the Dataset

- What type of information was collected?
  - Position information from a motion camera
  - Time stamps for each data point

- What are the units for each sample?
  - Position information should be in meters.
  - Time stamps should be in seconds.

- What should the data look like?
  - Individual numbers/values. Data should be “textual” and not graphical or videos.
  - Information should be human readable (as opposed to binary machine readable).

writing-data.xlsx

Download "Writing Data" from Canvas. Open in Excel

# Open the File and Examine the Columns. Are they as we expect?

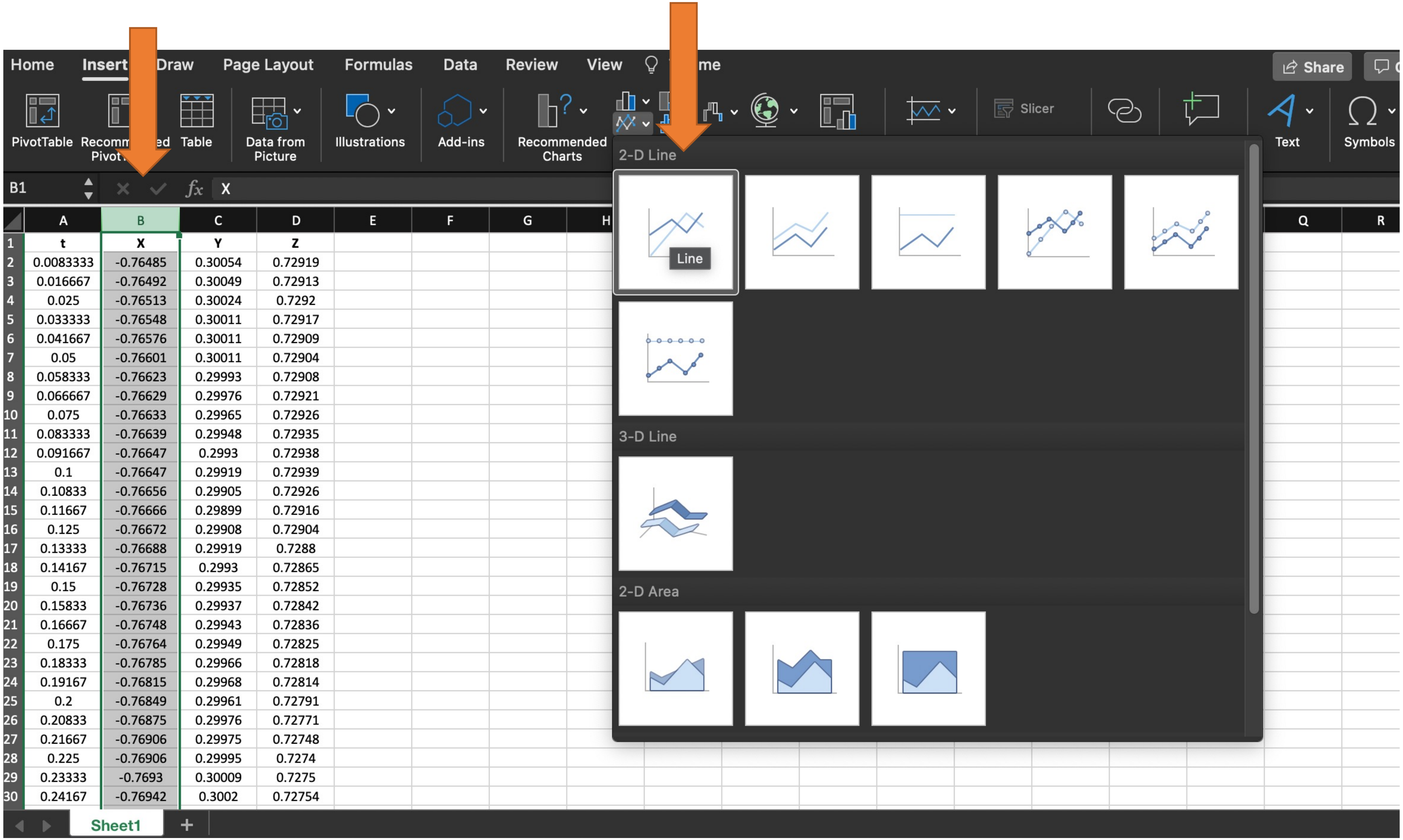
| Time (seconds) | X Position (meters) | Y Position (meters) | Z Position (meters) |
|---|---|---|---|
| **t** | **X** | **Y** | **Z** |
| 0.0083333 | -0.76485 | 0.30054 | 0.72919 |
| 0.016667 | -0.76492 | 0.30049 | 0.72913 |
| 0.025 | -0.76513 | 0.30024 | 0.7292 |
| 0.033333 | -0.76548 | 0.30011 | 0.72917 |
| 0.041667 | -0.76576 | 0.30011 | 0.72909 |
| 0.05 | -0.76601 | 0.30011 | 0.72904 |

Ok, so it’s motion information from someone’s hand. How can we visualize this?

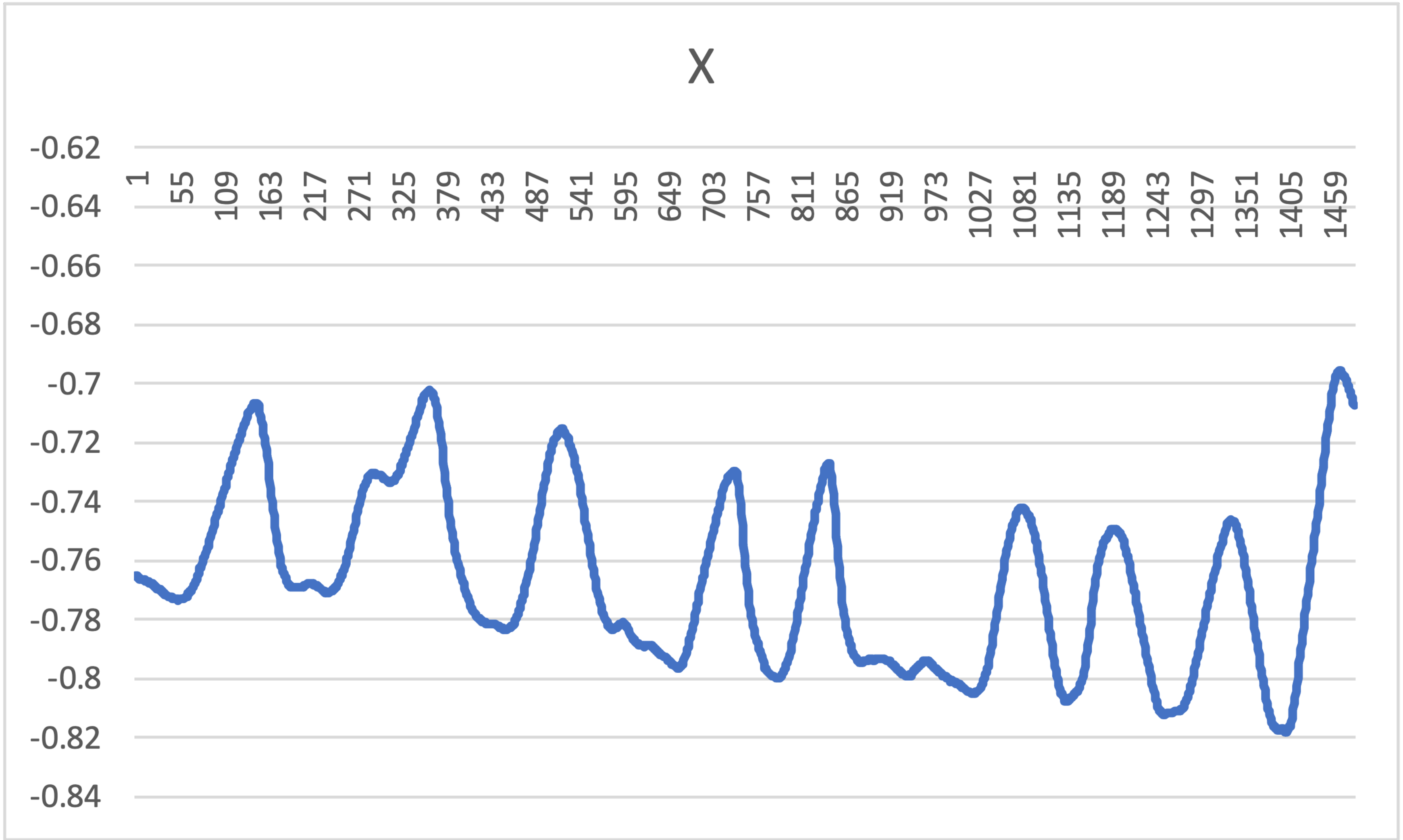
Create a “line plot” in Excel of a single column. See if that information is helpful.

## 1. Click/select the whole column

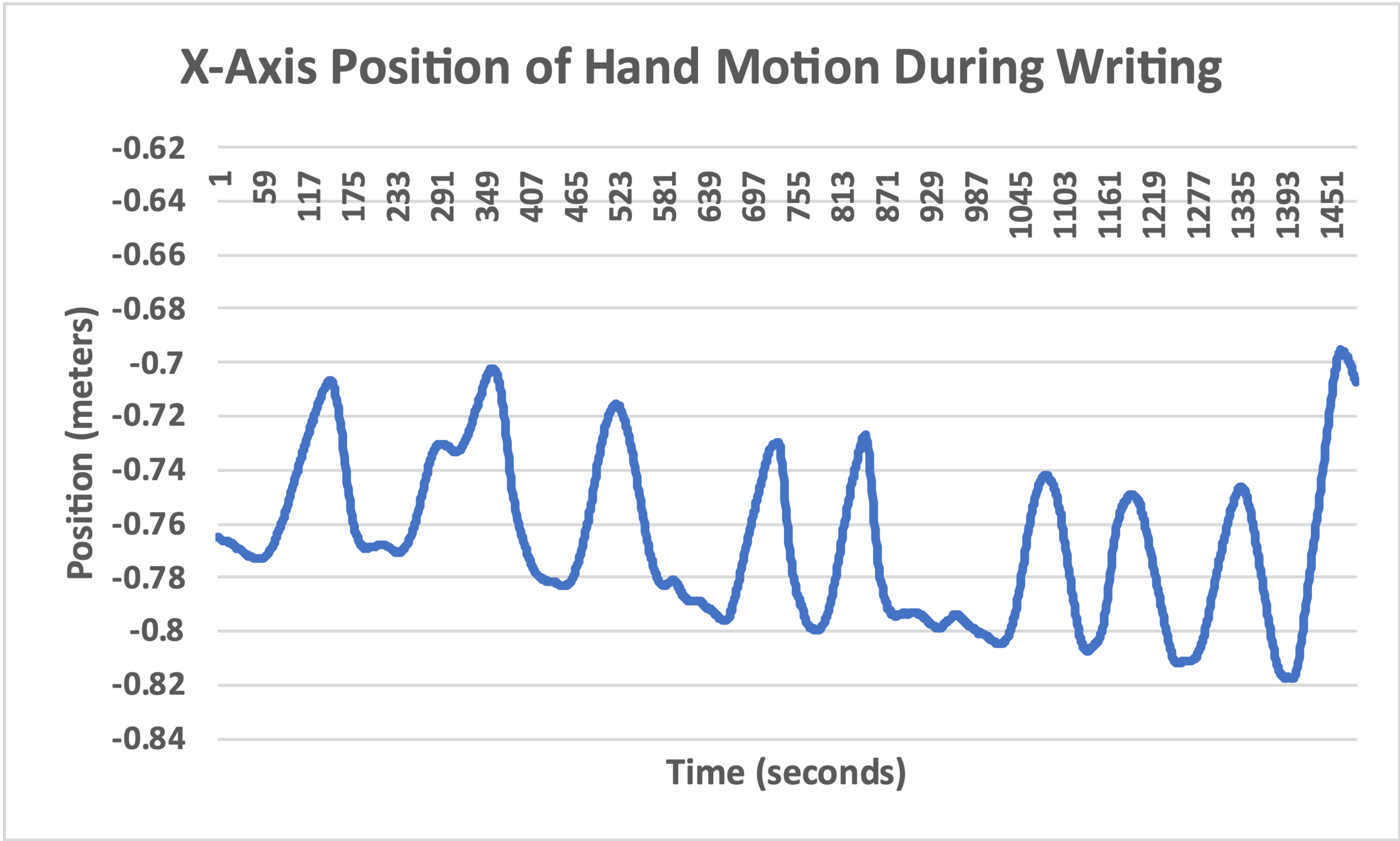
## 2. Under charts, select 2-D Line



# Here’s the result. Is this useful? Is it a “good” graph based upon the rules we discussed?

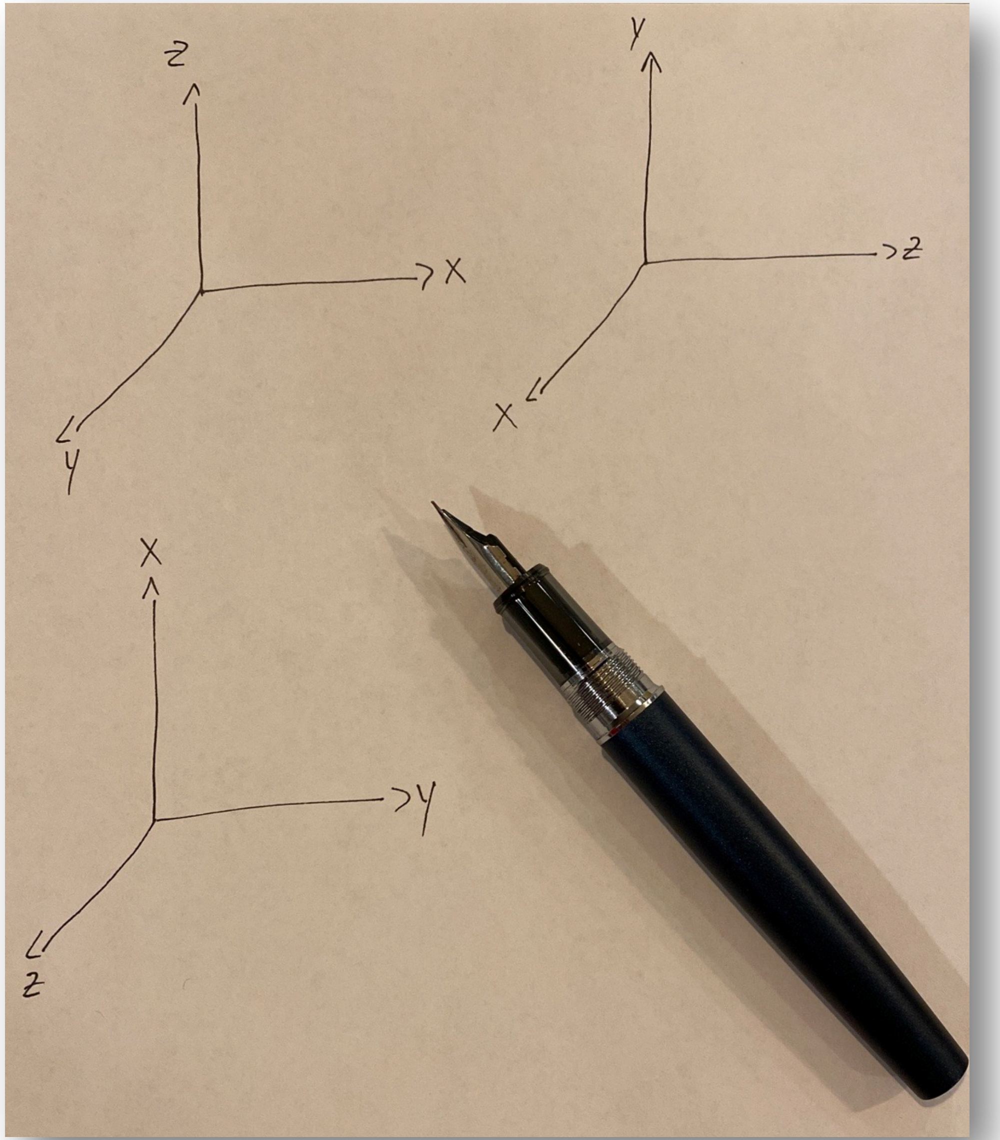


# Ok, this is better, but is the information useful? Can we tell what they were writing?



# Maybe we should re-think this...

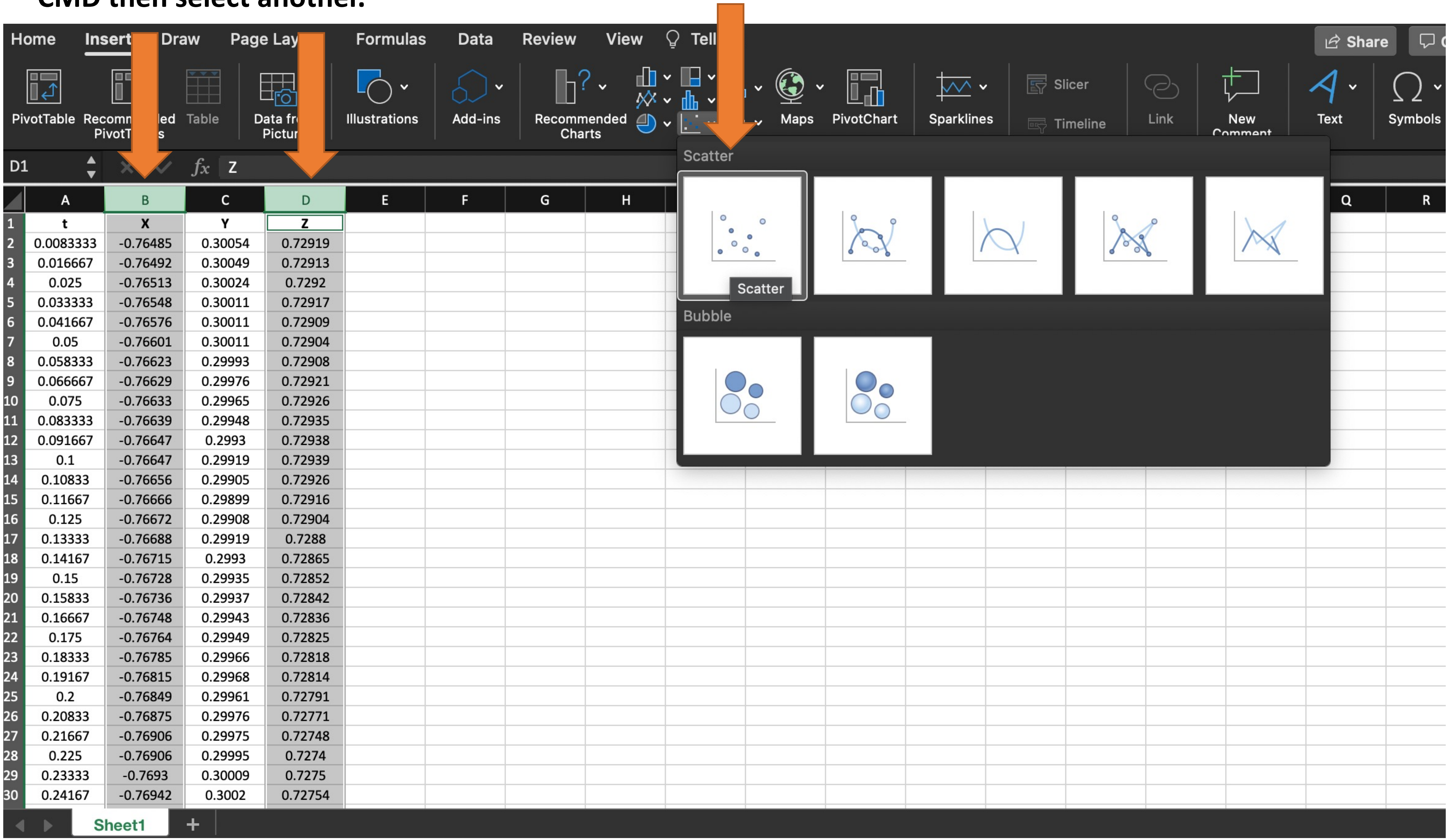
- Consider what was plotted. Only a single axis. If you plotted ”X”, what does that mean in the context of writing on a piece of paper? Up? Down? Left?
- Which orientation is correct?
- What tests (plots) might we try out to determine the orientation?



Make several Scatter plots by selection two columns at once. Discover the orientation of the data.

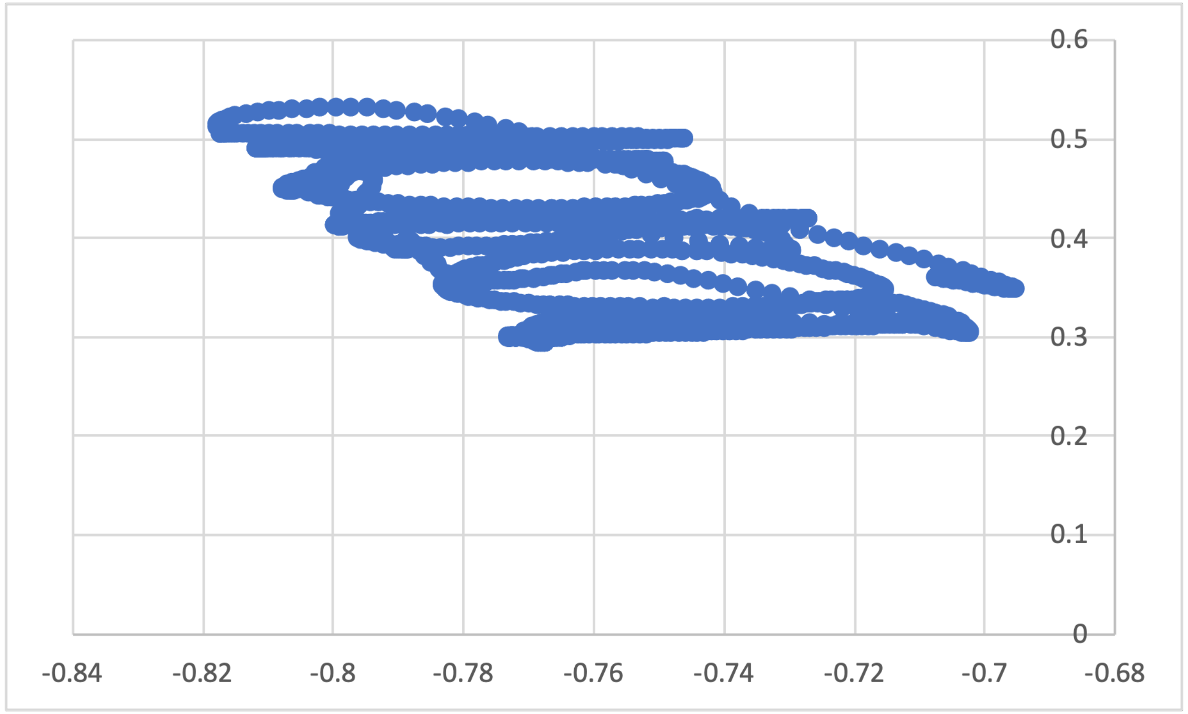
1. Select one column. Hold CRTL or CMD then select another.

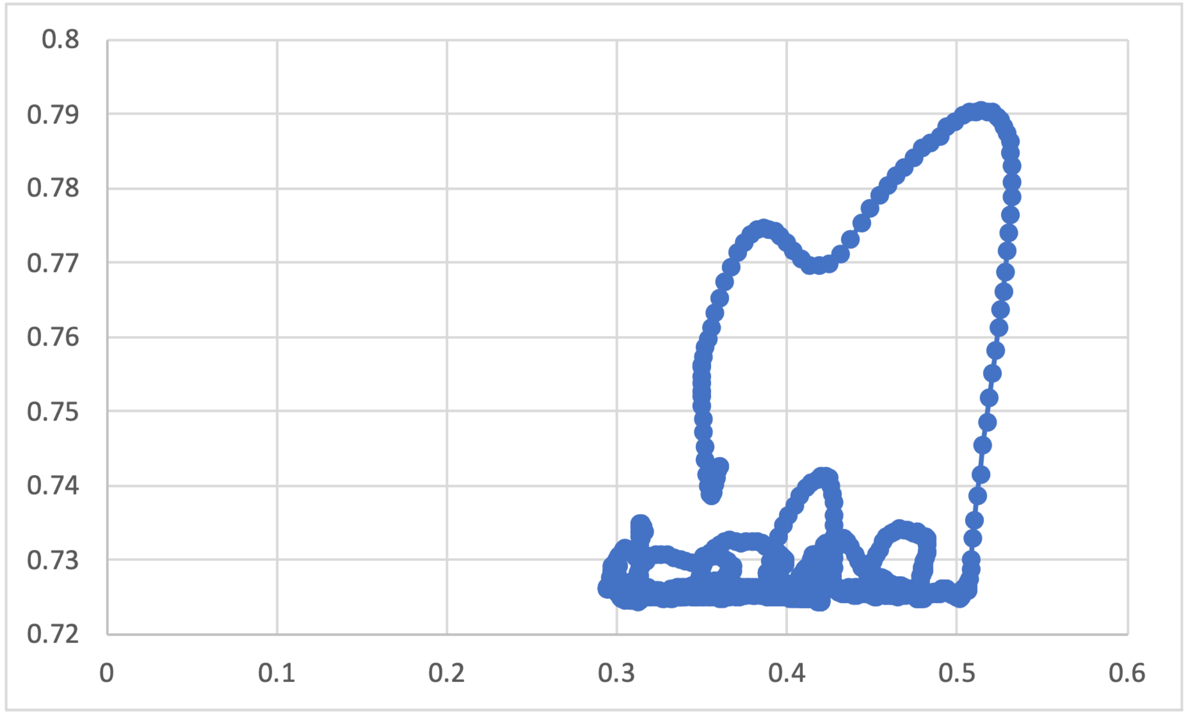
2. Create a scatter plot.




| | A | B | C | D |
|---|---|---|---|---|
| 1 | t | X | Y | Z |
| 2 | 0.0083333 | -0.76485 | 0.30054 | 0.72919 |
| 3 | 0.016667 | -0.76492 | 0.30049 | 0.72913 |
| 4 | 0.025 | -0.76513 | 0.30024 | 0.7292 |
| 5 | 0.033333 | -0.76548 | 0.30011 | 0.72917 |
| 6 | 0.041667 | -0.76576 | 0.30011 | 0.72909 |
| 7 | 0.05 | -0.76601 | 0.30011 | 0.72904 |
| 8 | 0.058333 | -0.76623 | 0.29993 | 0.72908 |
| 9 | 0.066667 | -0.76629 | 0.29976 | 0.72921 |
| 10 | 0.075 | -0.76633 | 0.29965 | 0.72926 |
| 11 | 0.083333 | -0.76639 | 0.29948 | 0.72935 |
| 12 | 0.091667 | -0.76647 | 0.2993 | 0.72938 |
| 13 | 0.1 | -0.76647 | 0.29919 | 0.72939 |
| 14 | 0.10833 | -0.76656 | 0.29905 | 0.72926 |
| 15 | 0.11667 | -0.76666 | 0.29899 | 0.72916 |
| 16 | 0.125 | -0.76672 | 0.29908 | 0.72904 |
| 17 | 0.13333 | -0.76688 | 0.29919 | 0.7288 |
| 18 | 0.14167 | -0.76715 | 0.2993 | 0.72865 |
| 19 | 0.15 | -0.76728 | 0.29935 | 0.72852 |
| 20 | 0.15833 | -0.76736 | 0.29937 | 0.72842 |
| 21 | 0.16667 | -0.76748 | 0.29943 | 0.72836 |
| 22 | 0.175 | -0.76764 | 0.29949 | 0.72825 |
| 23 | 0.18333 | -0.76785 | 0.29966 | 0.72818 |
| 24 | 0.19167 | -0.76815 | 0.29968 | 0.72814 |
| 25 | 0.2 | -0.76849 | 0.29961 | 0.72791 |
| 26 | 0.20833 | -0.76875 | 0.29976 | 0.72771 |
| 27 | 0.21667 | -0.76906 | 0.29975 | 0.72748 |
| 28 | 0.225 | -0.76906 | 0.29995 | 0.7274 |
| 29 | 0.23333 | -0.7693 | 0.30009 | 0.7275 |
| 30 | 0.24167 | -0.76942 | 0.3002 | 0.72754 |

# These plots are “correct” but do they reveal anything?

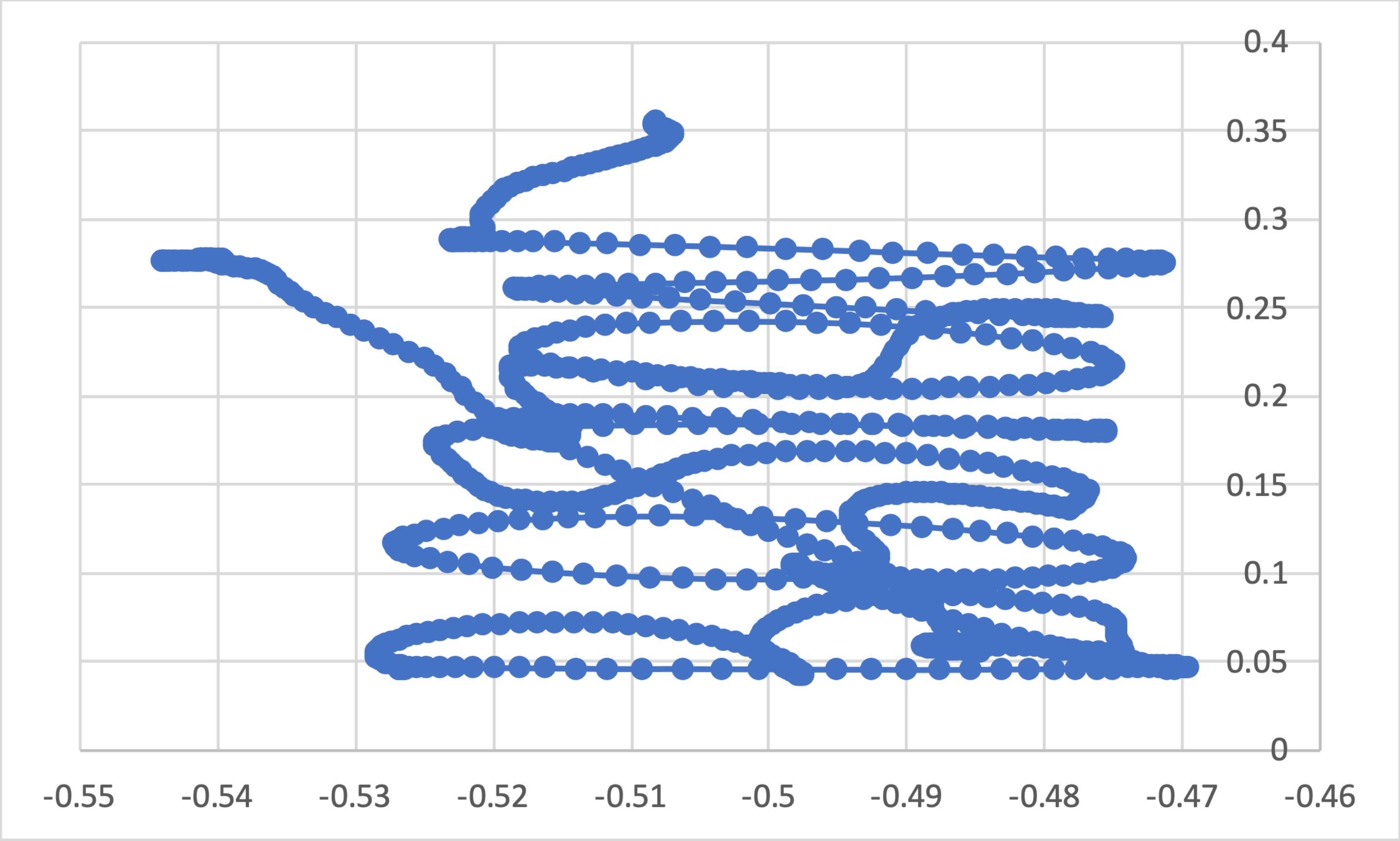


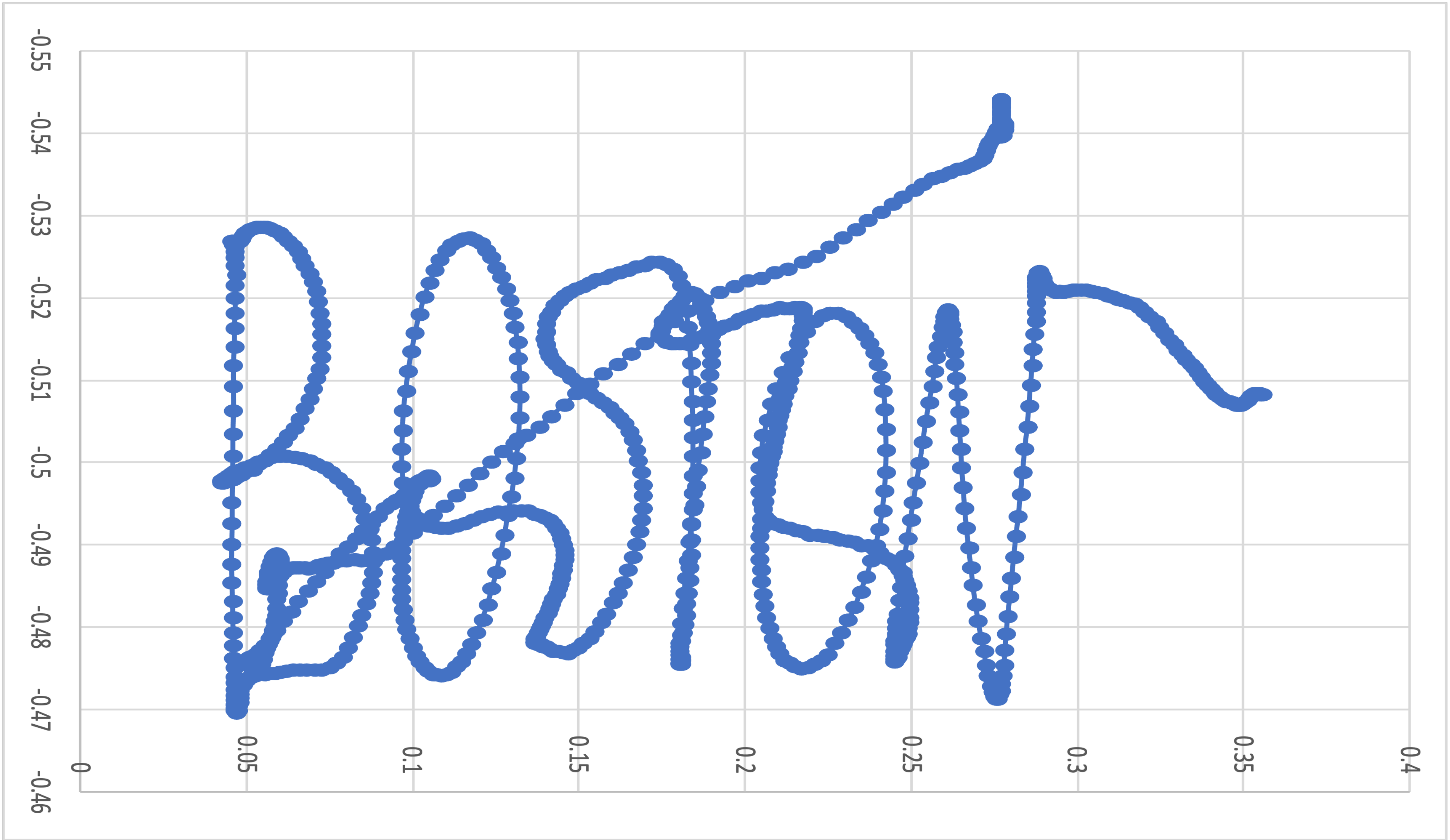
When analyzing a data set it’s important to look for specific artifacts and results.

This was a writing study. What were they tasked to write?

Look for the name of a city.



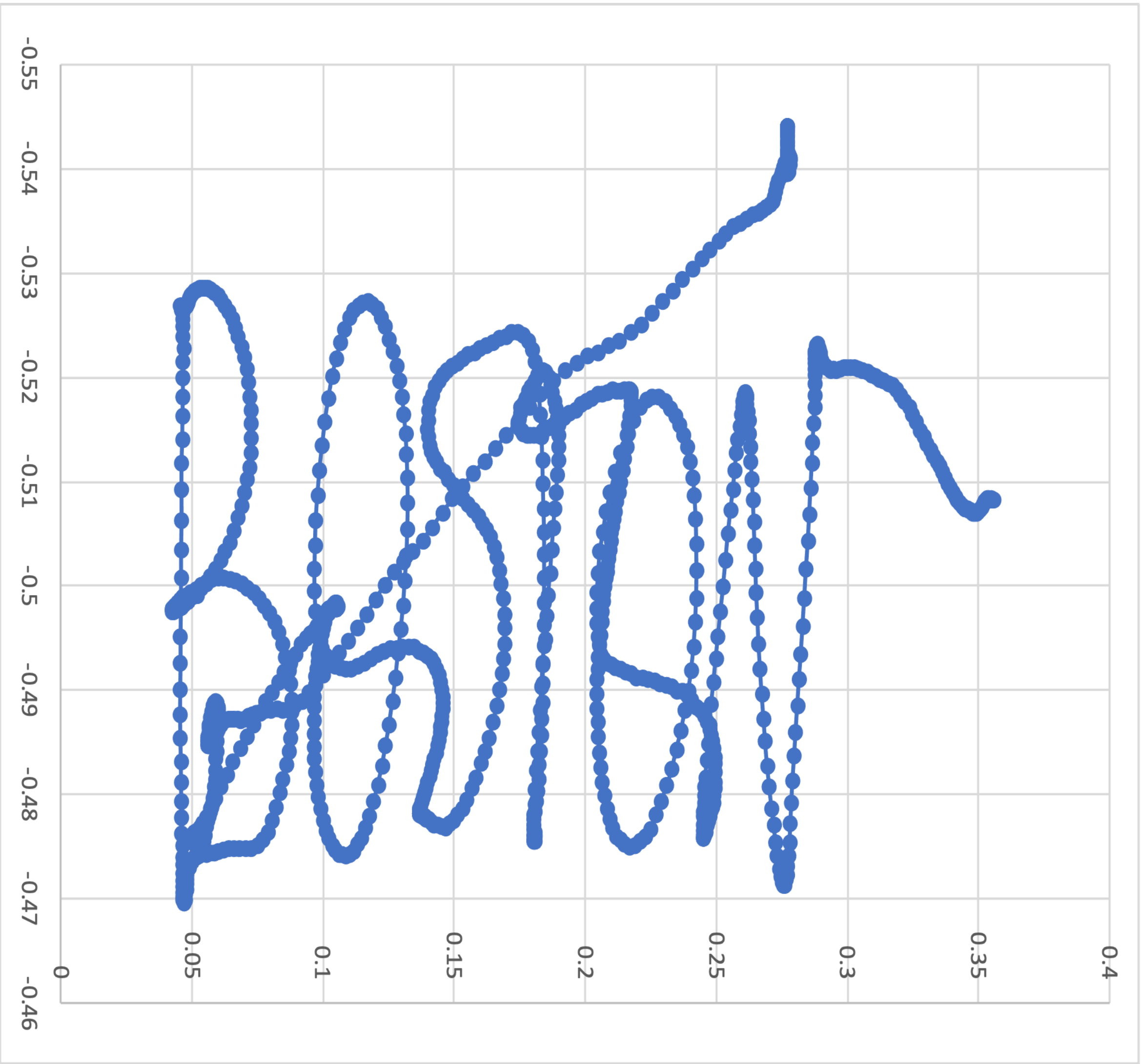
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
-0.55
-0.54
-0.53
-0.52
-0.51
-0.5
-0.49
-0.48
-0.47
-0.46


-0.55
-0.54
-0.53
-0.52
-0.51
-0.5
-0.49
-0.48
-0.47
-0.46
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4

**Boston**

# Reflecting on the Process

- Understand the format of your data and what it could show you.
- Examine different relationships to see if trends can be found and questions answered.
- All data has “noise” that can hide information. ”Boston” is not clearly written as the study tracked the person’s hand. We see that additional information and not just the pen marks.

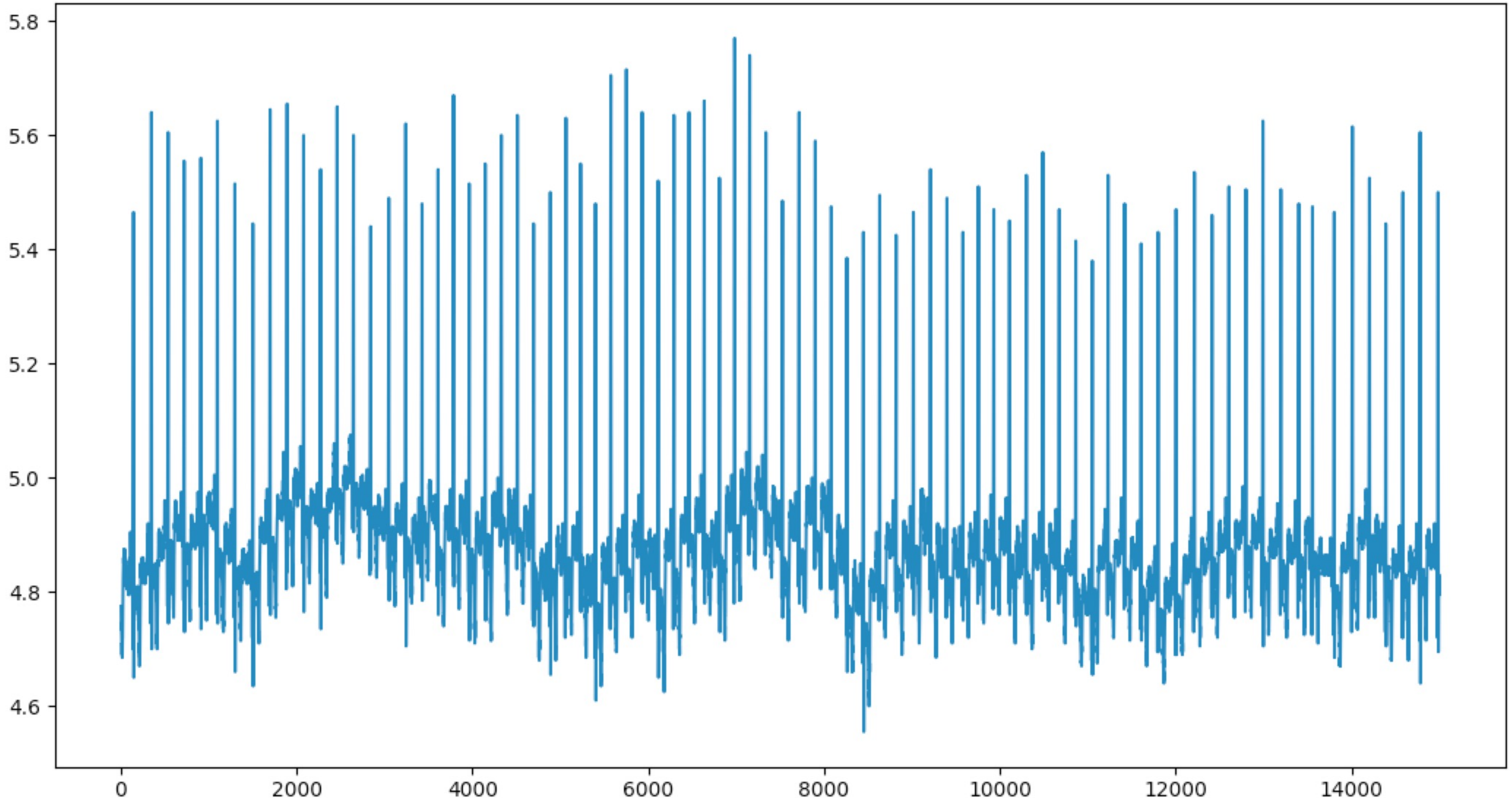


# Now for something different…

- The previous data set did not require any manipulation to “see” the results.

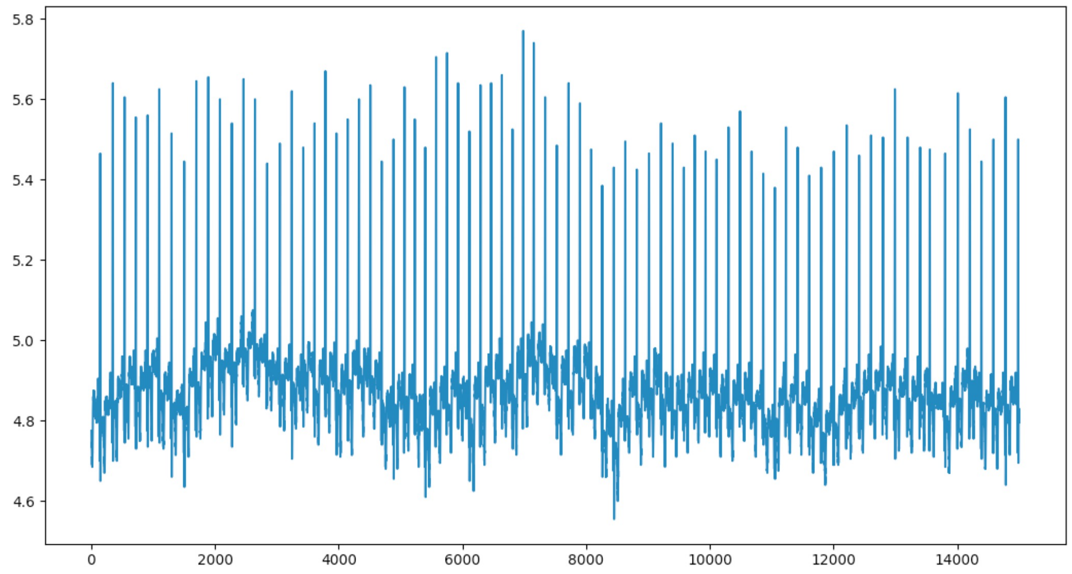
- Often data must be filtered, transformed, modified..etc to reveal information and/or to make it easier to process.

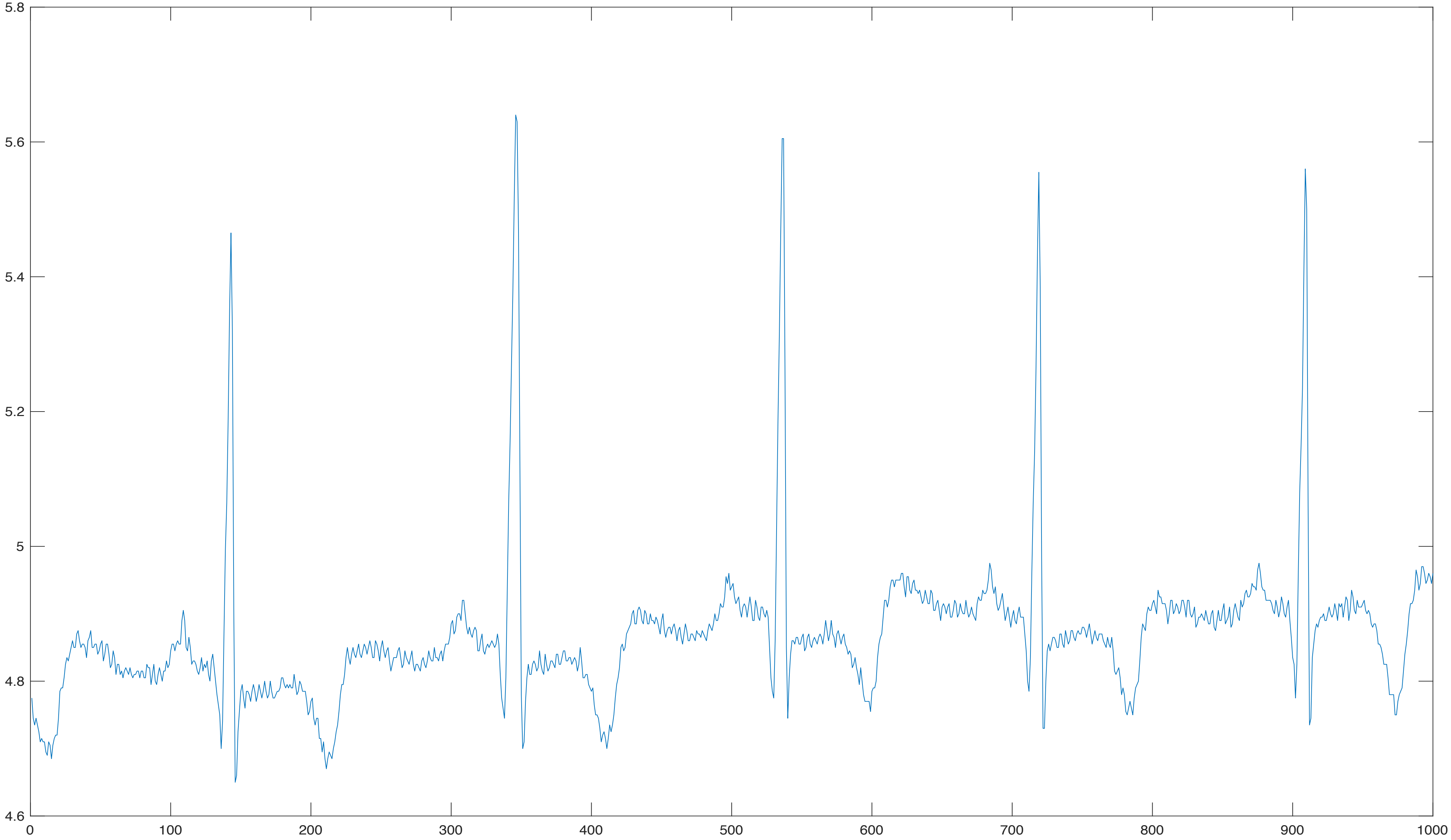
- The next dataset shows EKG readings. What might we wish to learn from it?


5.8
5.6
5.4
5.2
5.0
4.8
4.6
0
2000
4000
6000
8000
10000
12000
14000

# EKG Data

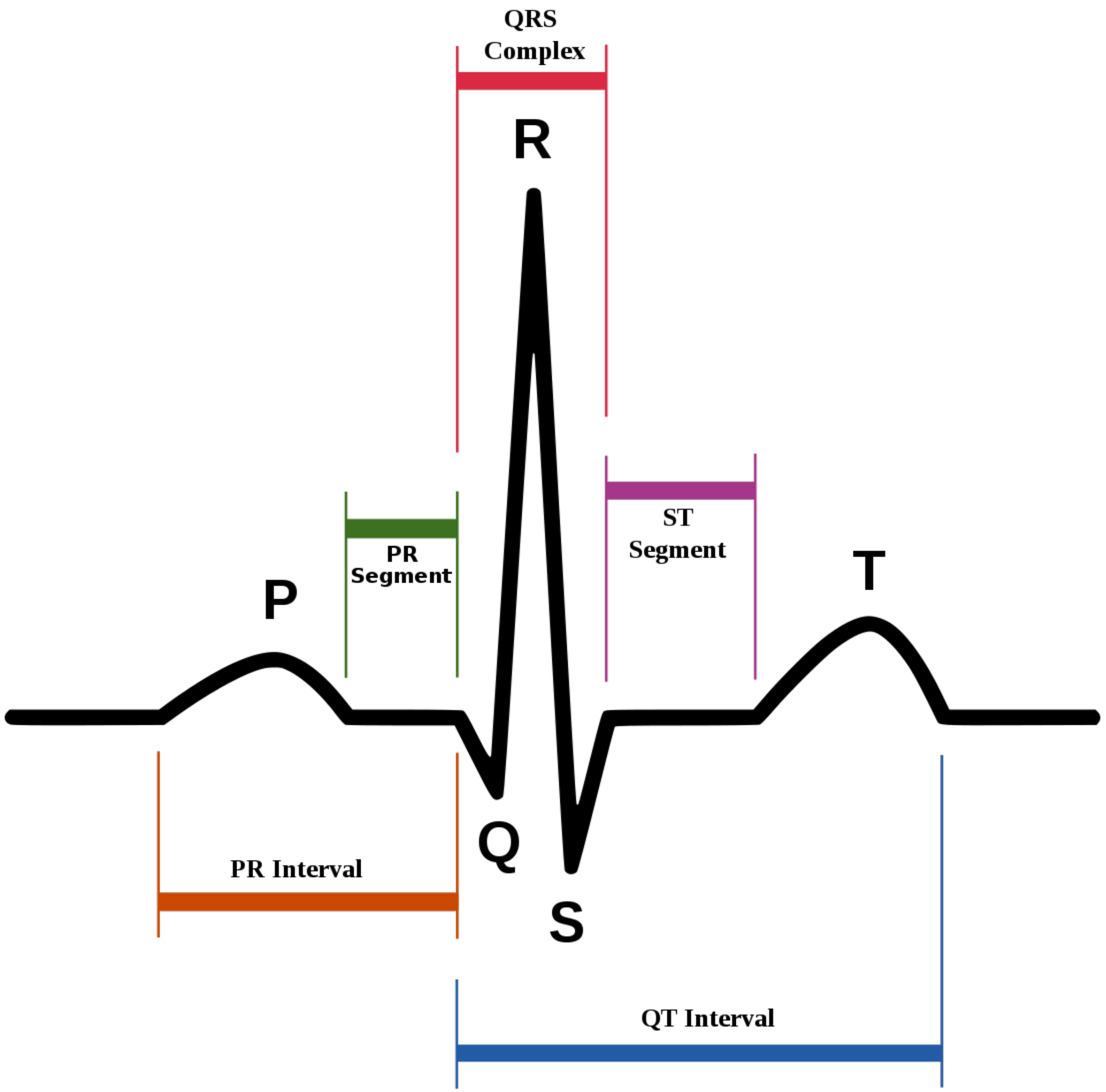
- What “kind” of information is being captured? Last data set was position over time.
- Why is there a ”rhythm” in the data? What physiological process is being captured?
- The recording is nearly 10s. What would it look like if we zoomed in?



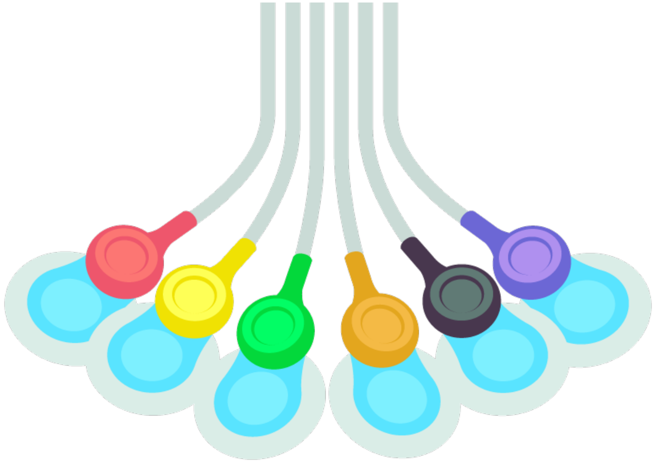

5.8
5.6
5.4
5.2
5
4.8
4.6
0
100
200
300
400
500
600
700
800
900
1000

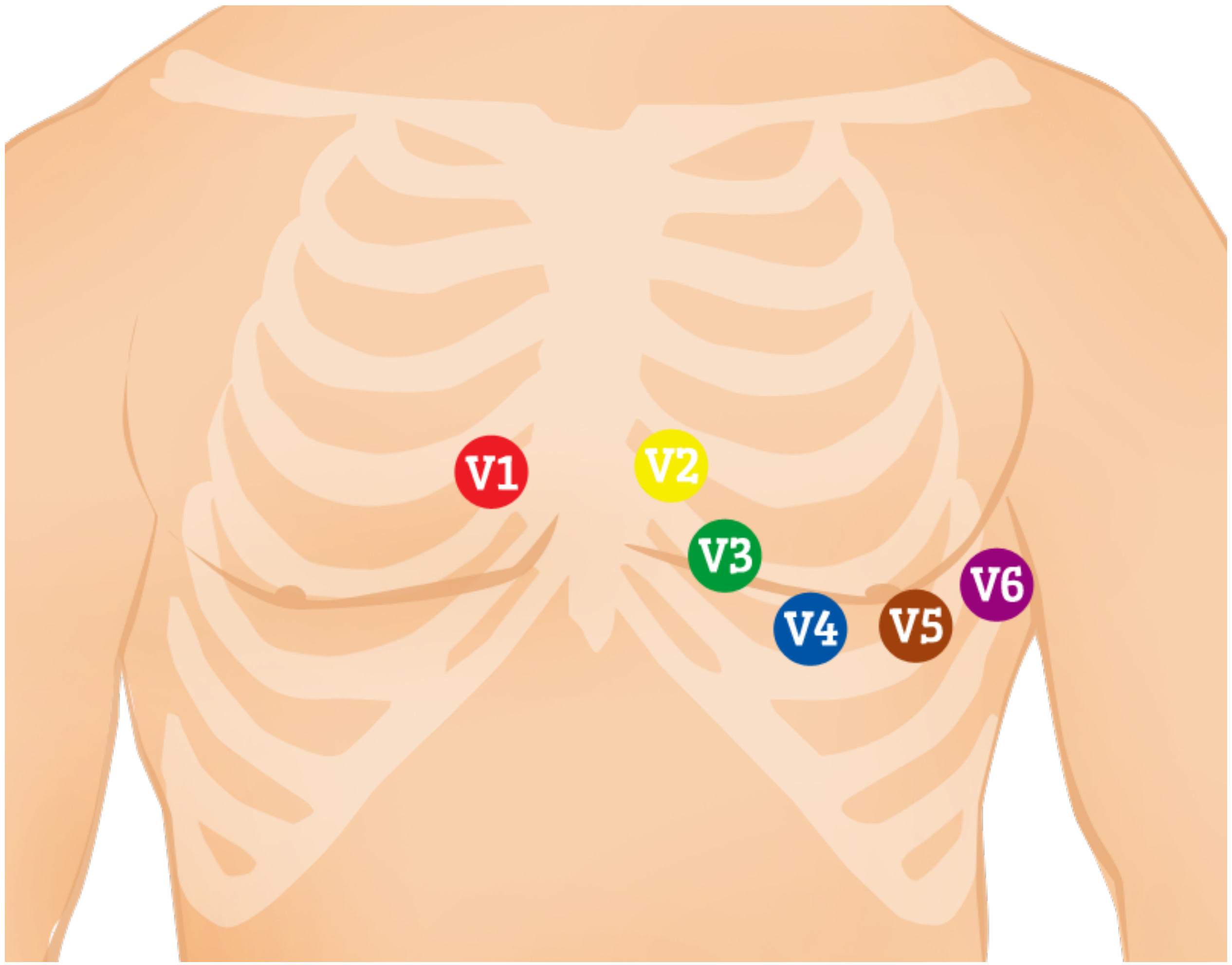
# QRS Wave

- What you are seeing are several QRS waves from heart electrical activity captured as voltage measurements.
- What kinds of ailments could we diagnose with this information?
- What would we need to calculate/analyze to identify those?



| Time (seconds) | MLII Lead (Volts) | V5 Lead (Volts) |
|---|---|---|
| Time (seconds) | MLII | V5 |
| 0 | 4.725 | 4.775 |
| 0.004 | 4.735 | 4.745 |
| 0.008 | 4.725 | 4.735 |
| 0.012 | 4.715 | 4.745 |
| 0.016 | 4.72 | 4.735 |
| 0.02 | 4.705 | 4.725 |
| 0.024 | 4.7 | 4.71 |

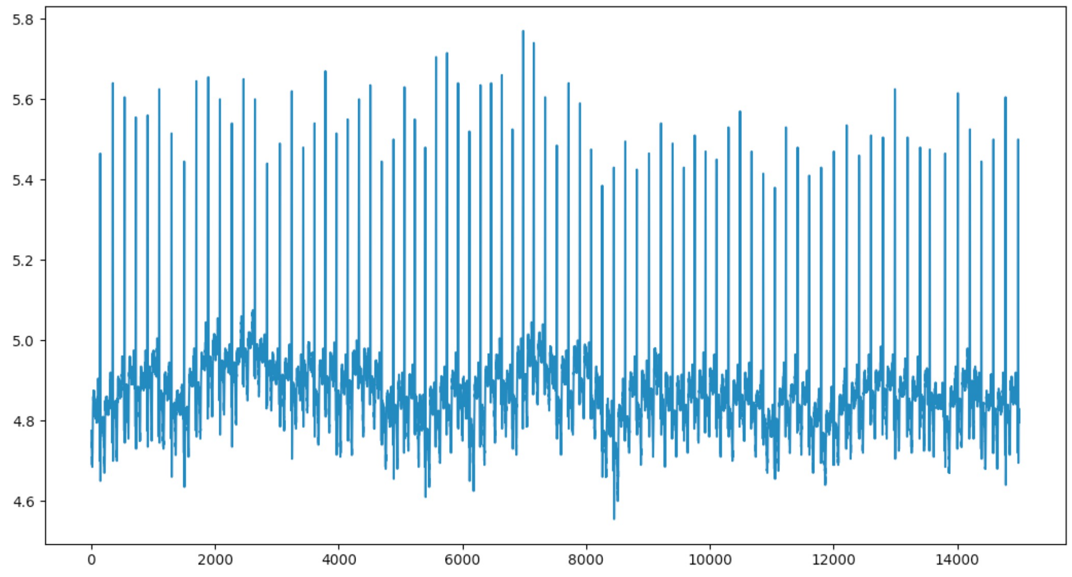




# Manipulating and Analyzing Data

- A basic cardiac measurement is determining Beats Per Minute. From there, can diagnose arrythmia and heart rate variability issues.

- Would be easy if we find each “peak” in the QRS wave. But the data is noisy, how can be make the determination easier?

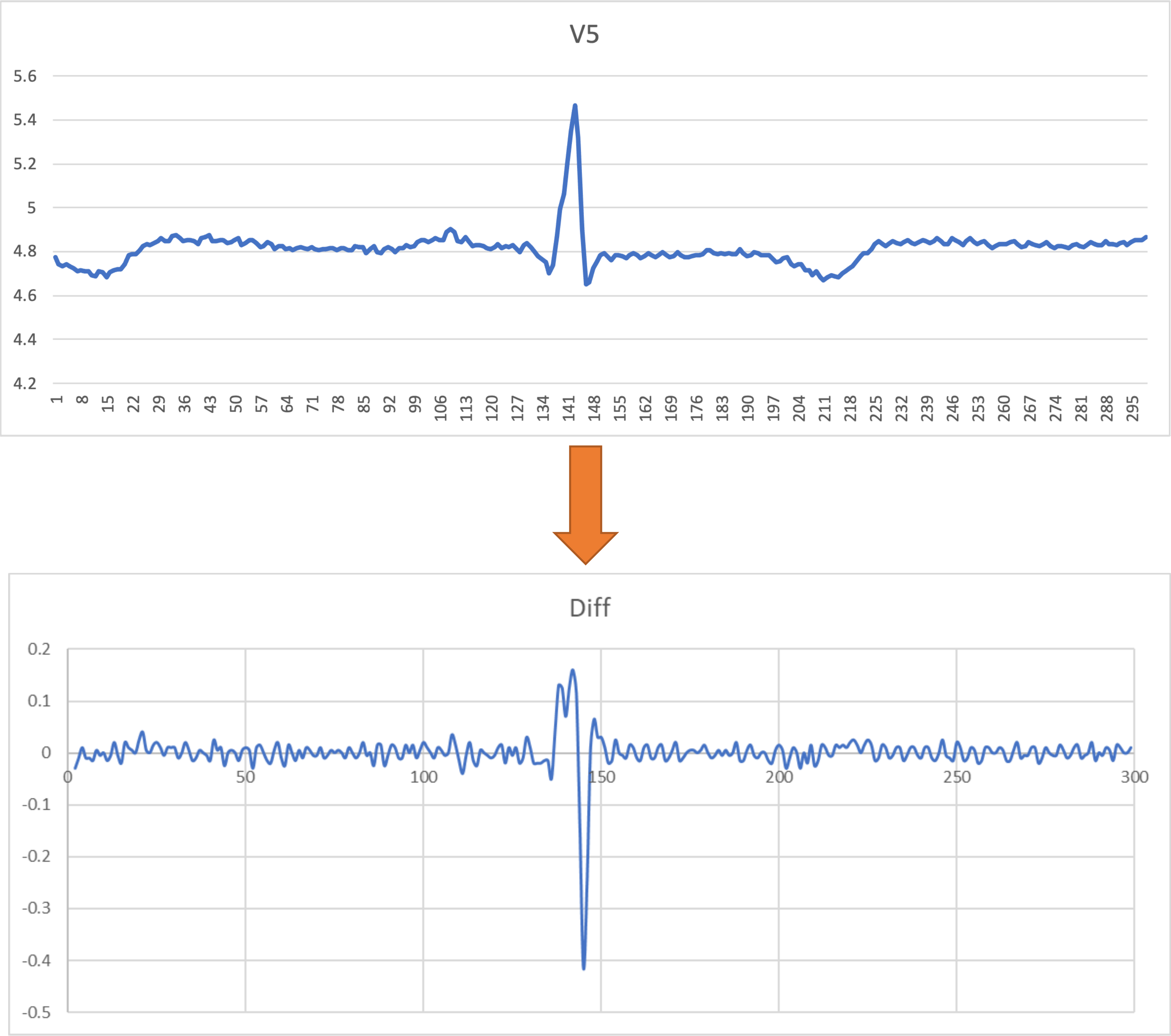


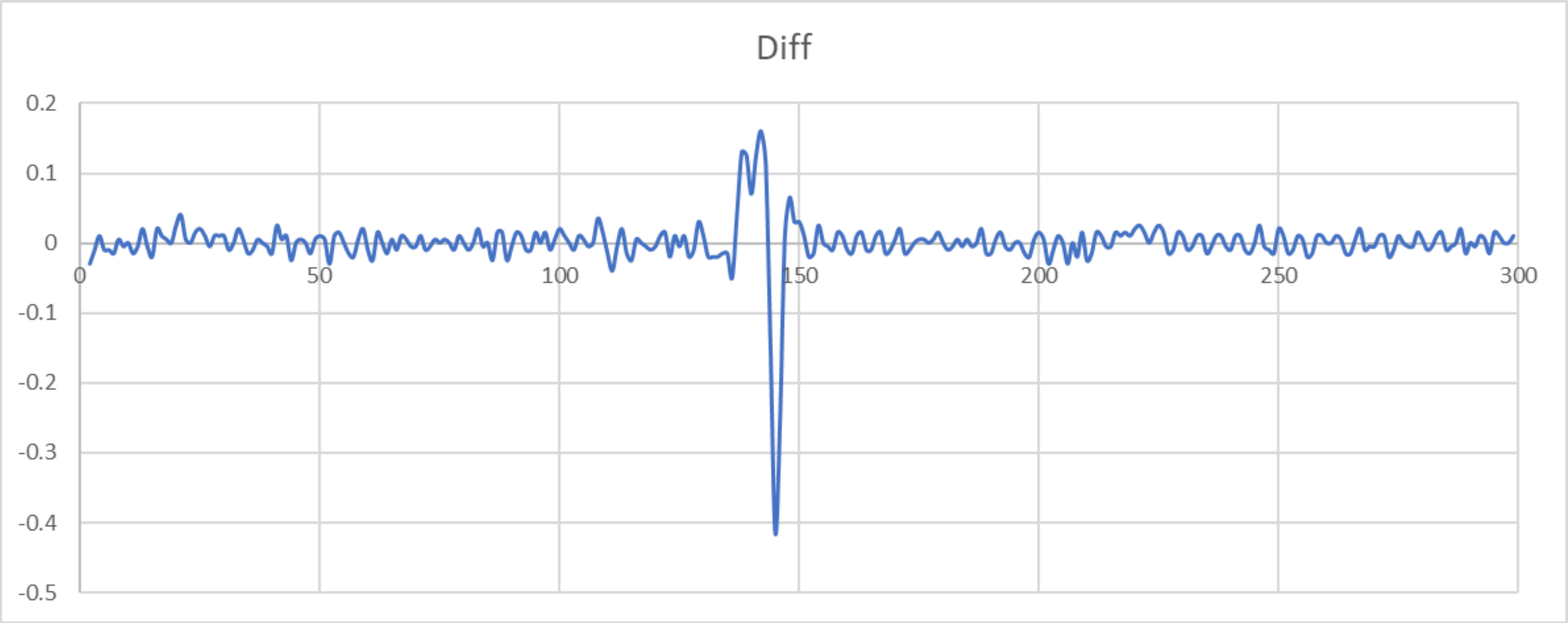
# Pan-Thompkins Algorithm for QRS Detection

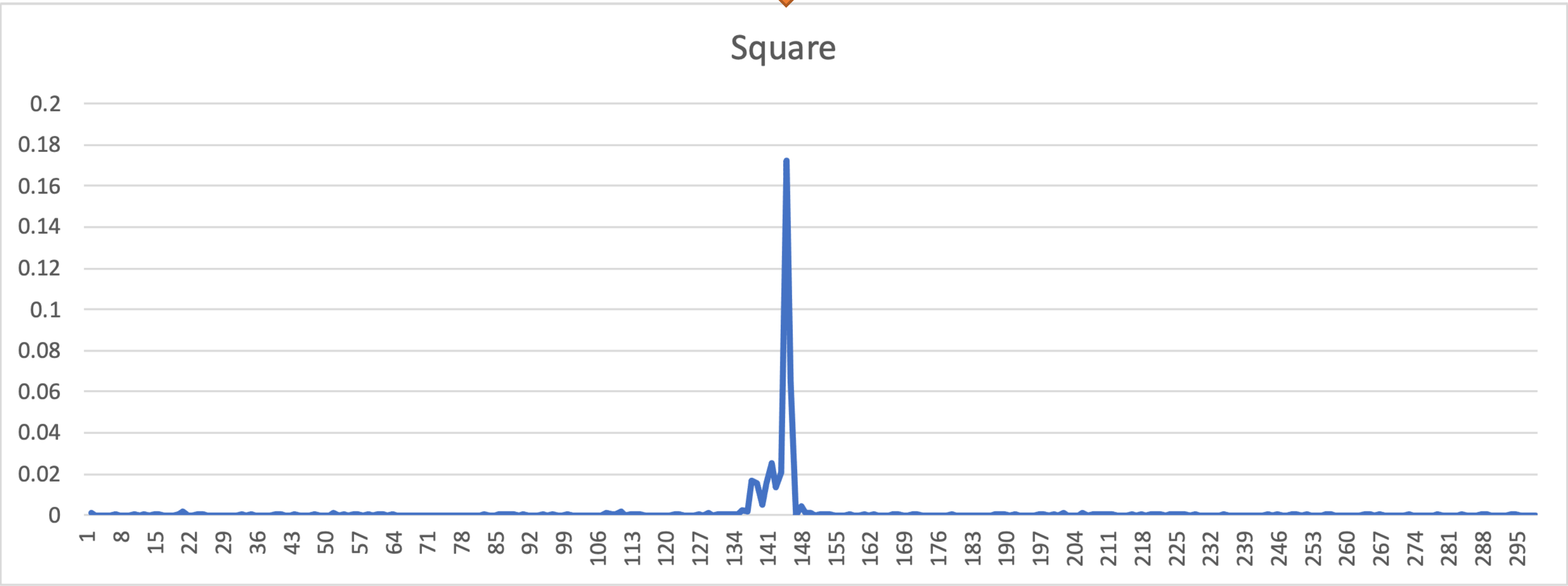
- To detect heart beats, we will follow a standard algorithm to isolate the "energy" from each heart beat. Will create a "pipeline" for processing the data.

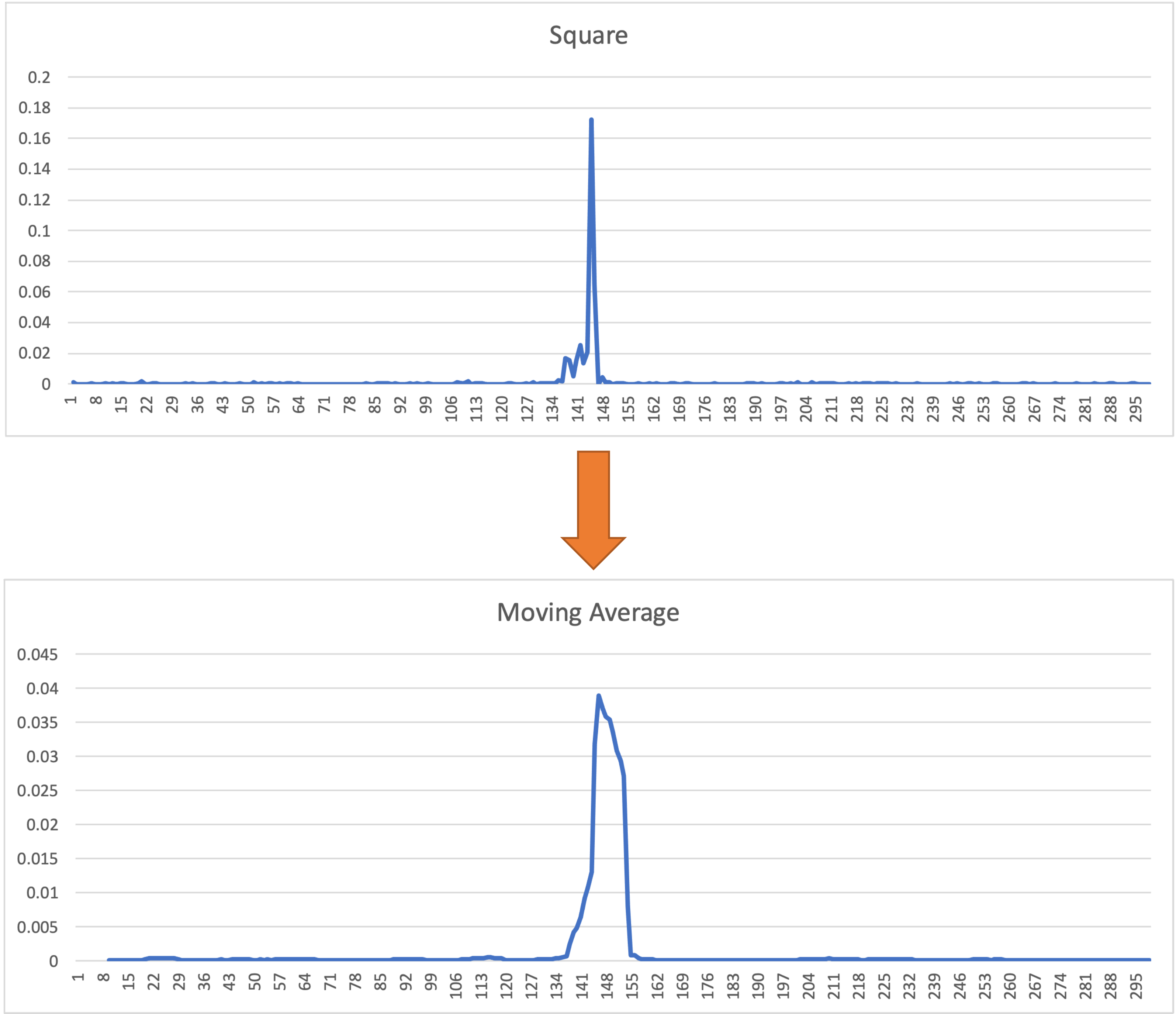
- Each signal is an array/vector of sequential values from [1-n]: $x_1, x_2, \ldots, x_n$. Each operation will result in a new array/vector.

- Three major operations to be performed:
  - Diff: take the difference between sequential points $y_1 = x_1 - x_2$
  - Square: take the result of Diff and perform ^2 operation $z_1 = y_1 * y_1$
  - Moving average: take the result from Diff and average some M values $Q = \sum_1^M z_i$

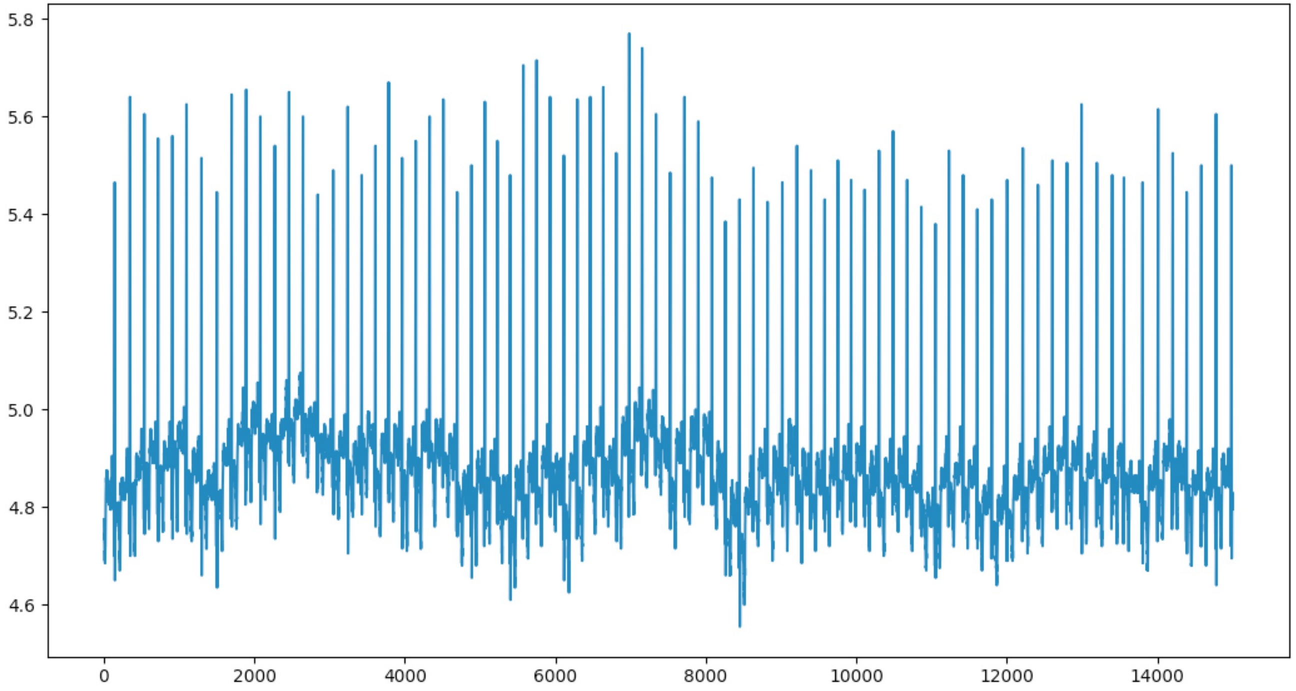

V5
5.6
5.4
5.2
5
4.8
4.6
4.4
4.2
1 8 15 22 29 36 43 50 57 64 71 78 85 92 99 106 113 120 127 134 141 148 155 162 169 176 183 190 197 204 211 218 225 232 239 246 253 260 267 274 281 288 295
Diff
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
-0.5
0 50 100 150 200 250 300


Diff
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
-0.5
0
50
100
150
200
250
300

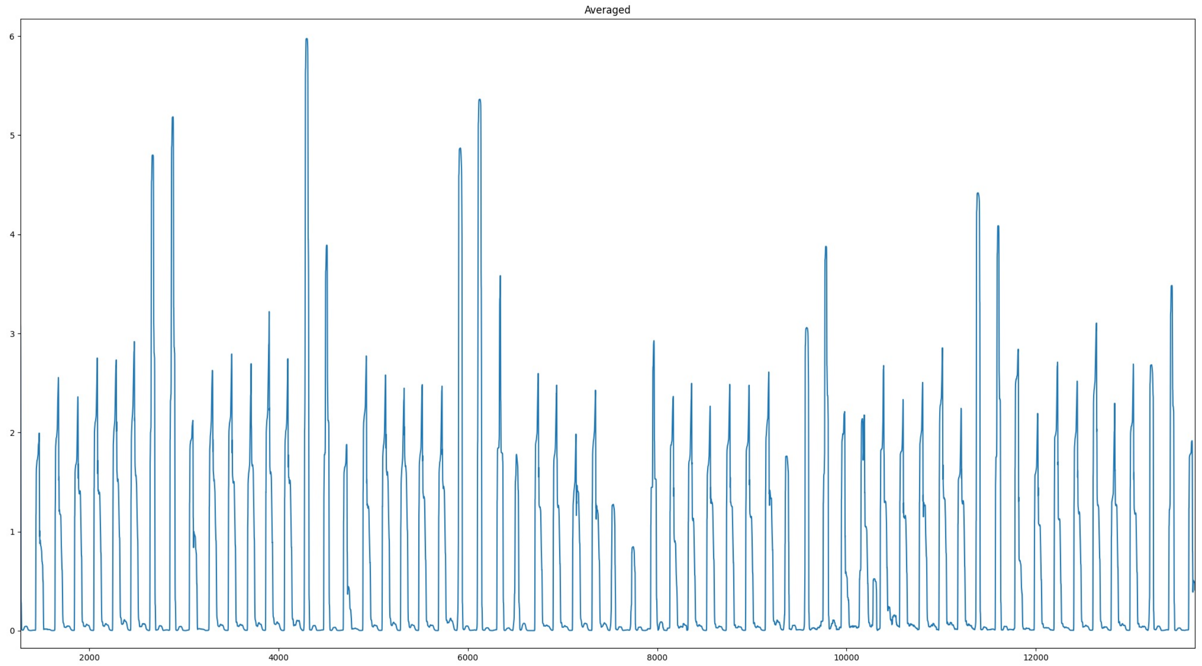

Square
0.2
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
1
8
15
22
29
36
43
50
57
64
71
78
85
92
99
106
113
120
127
134
141
148
155
162
169
176
183
190
197
204
211
218
225
232
239
246
253
260
267
274
281
288
295


Square
Moving Average

# Algorithm “Cleans up” the data. Which is easier to detect heart beats in?



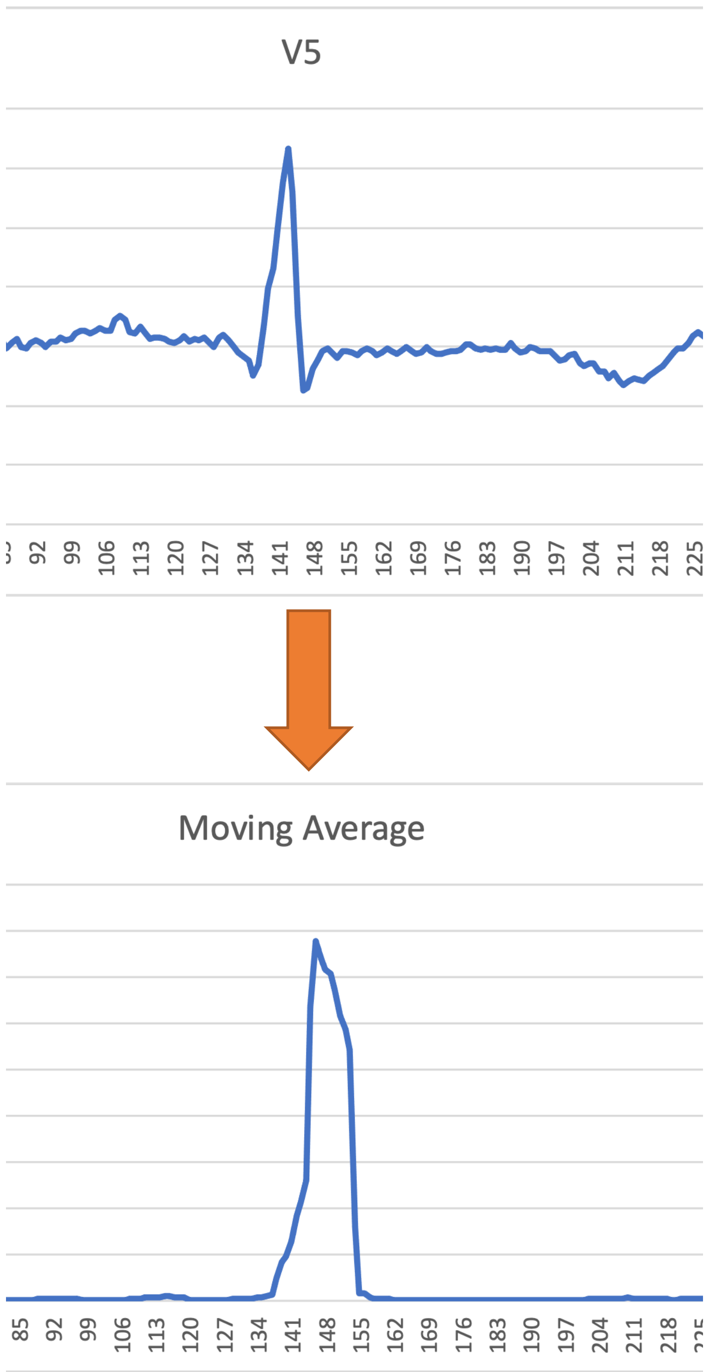

Raw ECG Data (before processing)




Post-processing and Filtered

# Some important items to notice

- The “peak” is now more prominent, relative to the signal “floor”, so it is easier to “see”.
- Signal is generally “0” when there is no peak nearby. This makes detection easier.
- Minor noise in signal is removed. However, signal is time delayed due to processing (average filter)



ekg-data.xlsx

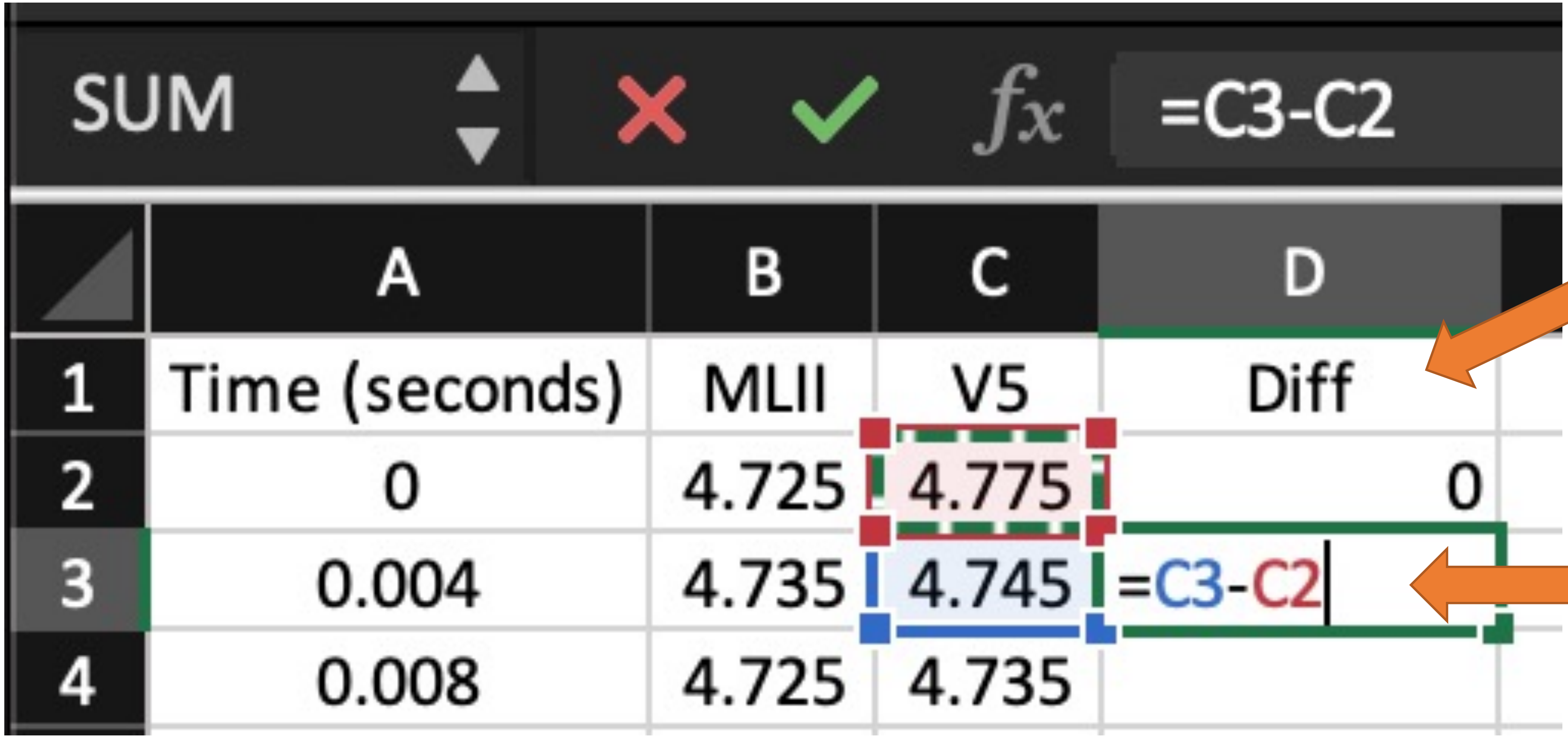
Download “ekg-data” from Canvas. Open in Excel

# Signal Analysis in Excel

- Leave the original data alone and process each “signal” as a new column. Use the basic +/* operators in Excel and the average() function.
- Write the equation once and then copy it down the excel column. This will propagate the equation into each cell.
- Some initial values may be unknown or deleted. Can’t have a result from a 10 data point average after the second result.

| Time (seconds) | MLII | V5 | **Diff** | **Square** | **Moving Average** |
|---|---|---|---|---|---|
| 0 | 4.725 | 4.775 | | | |
| 0.004 | 4.735 | 4.745 | -0.03 | 0.0009 | |
| 0.008 | 4.725 | 4.735 | -0.01 | 1E-04 | |
| 0.012 | 4.715 | 4.745 | 0.01 | 1E-04 | |
| 0.016 | 4.72 | 4.735 | -0.01 | 1E-04 | |
| 0.02 | 4.705 | 4.725 | -0.01 | 0.0001 | |
| 0.024 | 4.7 | 4.71 | -0.015 | 0.000225 | |
| 0.028 | 4.69 | 4.715 | 0.005 | 2.5E-05 | |
| 0.032 | 4.71 | 4.71 | -0.005 | 2.5E-05 | |
| 0.036 | 4.71 | 4.71 | 0 | 0 | 0.000175 |
| 0.04 | 4.695 | 4.695 | -0.015 | 0.000225 | 1E-04 |
| 0.044 | 4.69 | 4.69 | -0.005 | 2.5E-05 | 9.16667E-05 |
| 0.048 | 4.71 | 4.71 | 0.02 | 0.0004 | 0.000125 |
| 0.052 | 4.695 | 4.705 | -0.005 | 2.5E-05 | 0.000116667 |
| 0.056 | 4.705 | 4.685 | -0.02 | 0.0004 | 0.00015 |

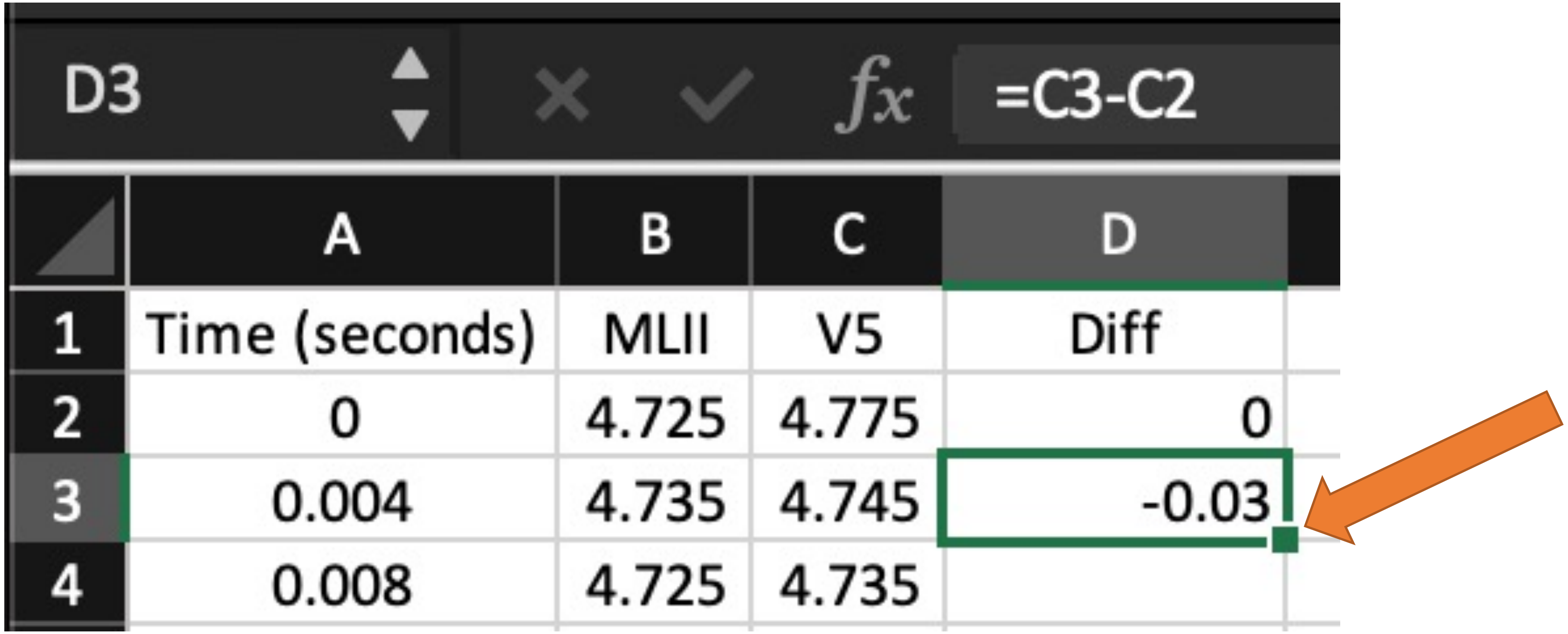
# Creating the First Diff Entry




**1. Give the column a name. Fill in the first value as 0**

**2. In the cell, type "=" and then click the cell in the V5 column. Type "-". Then click the next cell. Press enter when done.**

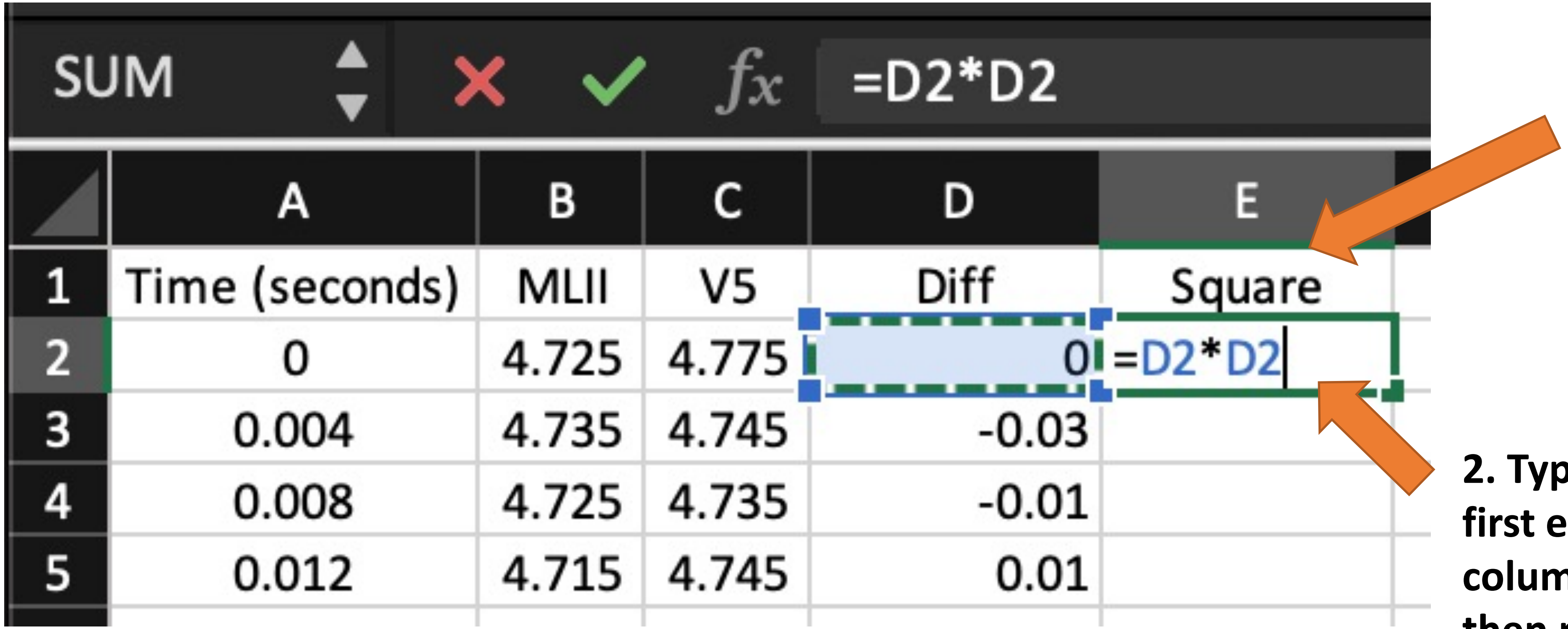
# When Done, double click the bottom right of cell to copy down the entire column.



D3 =C3-C2

| | A | B | C | D |
|---|---|---|---|---|
| 1 | Time (seconds) | MLII | V5 | Diff |
| 2 | 0 | 4.725 | 4.775 | 0 |
| 3 | 0.004 | 4.735 | 4.745 | -0.03 |
| 4 | 0.008 | 4.725 | 4.735 | |

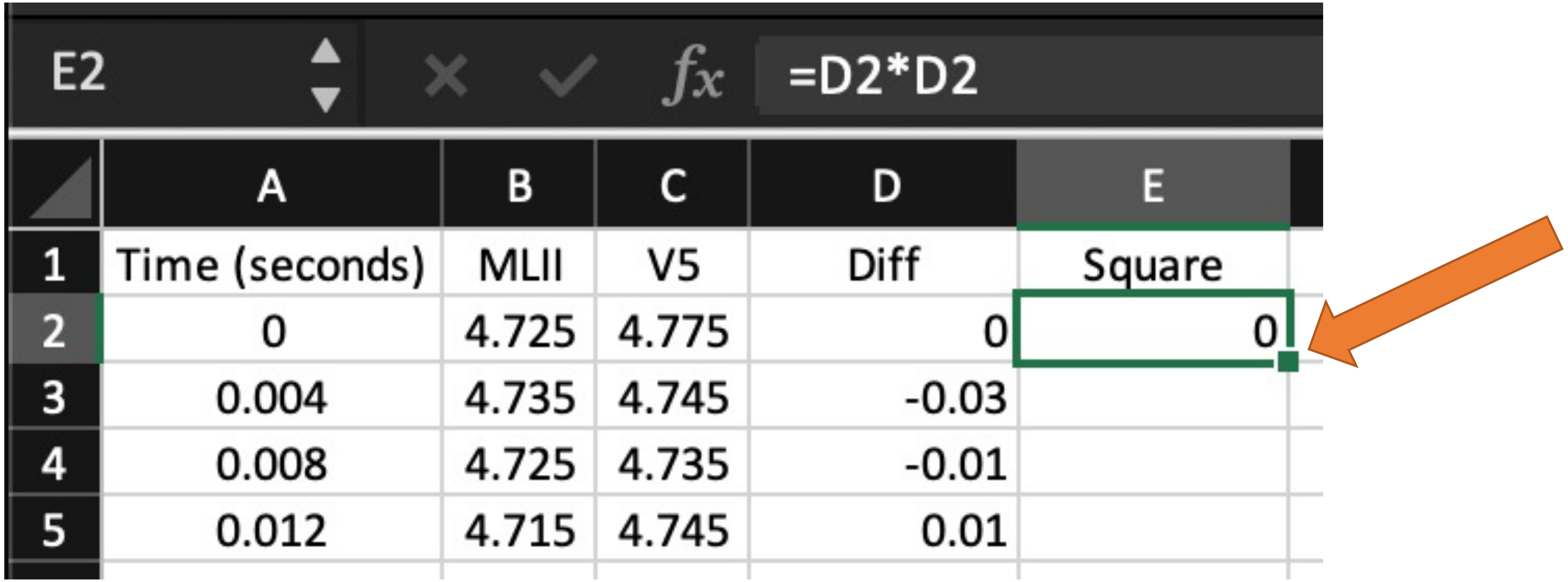
# Creating the Square Column




**1. Give the column a name.**

**2. Type "=". Click the first entry in the Diff column. Type "*" and then re-select the entry again. Press enter**

# When Done, double click the bottom right of cell to copy down the entire column.



E2 =D2*D2

| | A | B | C | D | E |
|---|---|---|---|---|---|
| 1 | Time (seconds) | MLII | V5 | Diff | Square |
| 2 | 0 | 4.725 | 4.775 | 0 | 0 |
| 3 | 0.004 | 4.735 | 4.745 | -0.03 | |
| 4 | 0.008 | 4.725 | 4.735 | -0.01 | |
| 5 | 0.012 | 4.715 | 4.745 | 0.01 | |

# Make a Moving Average of 10pts. Can be adjusted.

SUM | =average(E2:E12)

| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| 1 | Time (seconds) | MLII | V5 | Diff | Square | Moving Avg |
| 2 | 0 | 4.725 | 4.775 | 0 | 0 | |
| 3 | 0.004 | 4.735 | 4.745 | -0.03 | 0.0009 | |
| 4 | 0.008 | 4.725 | 4.735 | -0.01 | 1E-04 | |
| 5 | 0.012 | 4.715 | 4.745 | 0.01 | 1E-04 | |
| 6 | 0.016 | 4.72 | 4.735 | -0.01 | 1E-04 | |
| 7 | 0.02 | 4.705 | 4.725 | -0.01 | 0.0001 | |
| 8 | 0.024 | 4.7 | 4.71 | -0.015 | 0.000225 | |
| 9 | 0.028 | 4.69 | 4.715 | 0.005 | 2.5E-05 | |
| 10 | 0.032 | 4.71 | 4.71 | -0.005 | 2.5E-05 | |
| 11 | 0.036 | 4.71 | 4.71 | 0 | 0 | |
| 12 | 0.04 | 4.695 | 4.695 | -0.015 | 0.000225 | =average(E2:E12) |
| 13 | 0.044 | 4.69 | 4.69 | -0.005 | 2.5E-05 | |
| 14 | 0.048 | 4.71 | 4.71 | 0.02 | 0.0004 | |

**This time, when typing in the formula use the built-in average function.**

**Type "=average(" then use the cursor to select a range of data (about 10 pts). Then type ")" to close the formula. Press enter.**

**Expand the formula down the column.**

You should have something like this…

M24

| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| 1 | Time (seconds) | MLII | V5 | Diff | Square | Moving Avg |
| 2 | 0 | 4.725 | 4.775 | 0 | 0 | |
| 3 | 0.004 | 4.735 | 4.745 | -0.03 | 0.0009 | |
| 4 | 0.008 | 4.725 | 4.735 | -0.01 | 1E-04 | |
| 5 | 0.012 | 4.715 | 4.745 | 0.01 | 1E-04 | |
| 6 | 0.016 | 4.72 | 4.735 | -0.01 | 1E-04 | |
| 7 | 0.02 | 4.705 | 4.725 | -0.01 | 0.0001 | |
| 8 | 0.024 | 4.7 | 4.71 | -0.015 | 0.000225 | |
| 9 | 0.028 | 4.69 | 4.715 | 0.005 | 2.5E-05 | |
| 10 | 0.032 | 4.71 | 4.71 | -0.005 | 2.5E-05 | |
| 11 | 0.036 | 4.71 | 4.71 | 0 | 0 | |
| 12 | 0.04 | 4.695 | 4.695 | -0.015 | 0.0[illegible]25 | 0.00016364 |
| 13 | 0.044 | 4.69 | 4.69 | -0.005 | 2.5E-05 | 0.00016591 |
| 14 | 0.048 | 4.71 | 4.71 | 0.02 | 0.0004 | 0.00012045 |
| 15 | 0.052 | 4.695 | 4.705 | -0.005 | 2.5E-05 | 0.00011364 |
| 16 | 0.056 | 4.705 | 4.685 | -0.02 | 0.0004 | 0.00014091 |
| 17 | 0.06 | 4.7 | 4.705 | 0.02 | 0.0004 | 0.00016818 |
| 18 | 0.064 | 4.71 | 4.715 | 0.01 | 1E-04 | 0.00016818 |
| 19 | 0.068 | 4.71 | 4.72 | 0.005 | 2.5E-05 | 0.00015 |
| 20 | 0.072 | 4.7 | 4.72 | 0 | 0 | 0.00014773 |
| 21 | 0.076 | 4.705 | 4.745 | 0.025 | 0.000625 | 0.00020227 |
| 22 | 0.08 | 4.72 | 4.785 | 0.04 | 0.0016 | 0.00034773 |
| 23 | 0.084 | 4.735 | 4.79 | 0.005 | 2.5E-05 | 0.00032955 |
| 24 | 0.088 | 4.74 | 4.79 | 0 | 0 | 0.00032727 |
| 25 | 0.092 | 4.73 | 4.805 | 0.015 | 0.000225 | 0.00031136 |

# Work on ECG Dataset. Submit as Individual Before End of Class.

## Analyzing Heart Beat Data

Published Edit

Following the pipeline provided in class, perform analysis to **determine the number of beats** in the ekg-data set.

Deliverables:

- Upload your solution as an Excel document.
- Create a line plot showing the original EKG data. Provide proper axis labels and title. Upload plot as PDF.
- Create a line plot showing the resulting moving average plot. Provide proper axis labels and title. Upload plot as PDF.