

---

---

# Data Transformation and Organization

---

---

The slide features decorative horizontal lines: a thick teal line at the top, a thin teal line below it, and another thick teal line at the bottom. Two short, thick olive-green dashes are positioned horizontally on the slide, one to the left and one to the right of the center, below the main title.

# Goals

Remember transformation definitions

Calculate transformations by hand

Calculate transformations with Google Sheets

Make univariate plots with Google Sheets

Consider how transformations/organization/selection impact honest reporting

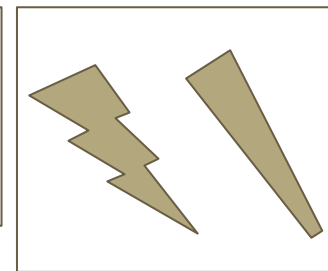
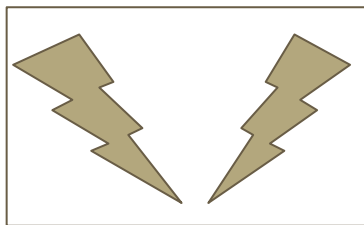
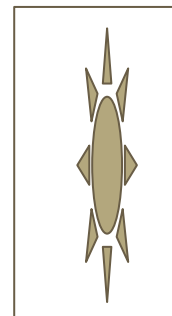
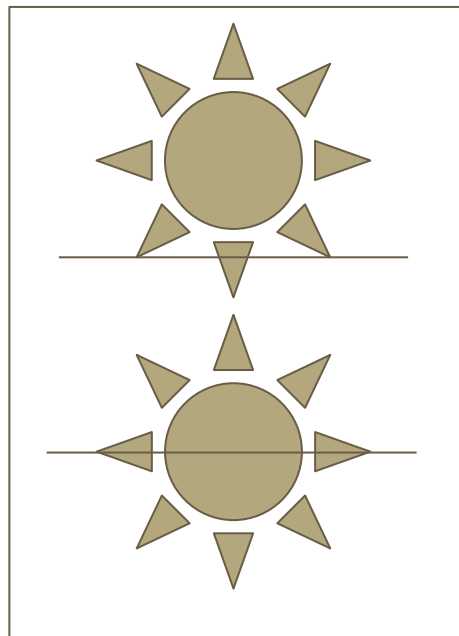
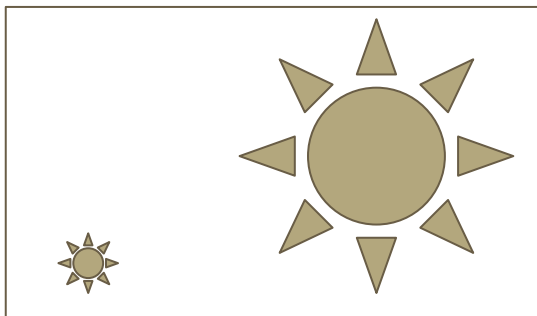
# Examples

- scaling
- centering
- normalization
- decimating
- shifting
- reflecting
- smoothing

## Transformation

- sorting
- selecting

## Organization



Data Dishonesty?

Data Dishonesty?

Data Dishonesty?

Data Dishonesty?

Data Dishonesty?

Data Dishonesty?



# Data

time,amplitude  
0,0  
1,101  
2,99  
3,102  
4,98  
5,105  
6,95  
7,110  
8,95  
9,104  
10,97  
11,102  
12,99  
13,101  
14,100  
15,99  
16,102  
17,98  
18,105  
19,95  
20,110  
21,95  
22,104  
23,97  
24,102  
25,99  
26,101  
27,100

> course-materials

> datasets

> transformation-data.csv

?

<https://tableconvert.com/markdown-to-excel>

# Steps

- download file
- open blank google sheets
- file > import
  - recommended: convert strings to numbers
  - recommended: select delimiter
- How many variables are there?
- How many observations are there?

# Automatic Plot

- Select variables
- Insert > chart
  - default: line
  
- Select one variable
- Insert > chart
  - default: line

## CAUTION

Why do they look the same?  
What about them is different?

## Other questions

- can you tell how many data points there are?
  - chart type: scatter

Google Sheets is simple, but limited

# Organize

- Select B
  - Group 1
    - Data > Sort Sheet > B: A-> Z
  - Group 2
    - Data > Sort Sheet > B: Z->A
  - Group 3
    - Data > Sort Range > B: A-> Z
  - Group 4
    - Data > Sort Range > B: Z->A

**Sorting**

- Select B
  - Data > Add Filter
    - drop down to select desired elements

**Filtering**



# Transform

- scaling → multiplication or division
- centering → subtracting out the mean
- normalization → adjusting the min and max
- decimating → dividing out a base number until data point is close to 1
- shifting → constant addition or subtraction (translation)
- reflecting → relative addition or subtraction
- smoothing → moving averages

# Transformations

Scale	multiplication or division	= cell * scalar
Mean		= AVERAGE(cell:cell)
Center	subtracting out the mean	= cell - \$Mean\$cell
Normalization to 1	scaling by absolute largest value	
Absolute	value regardless of sign	= ABS(cell)
Largest		=MAX(cell:cell)

# Transformations

Shift	add or subtract constant	=cell $\pm$ constant
Reflect	add or subtract relative	=cell $\pm$ 2*(\$reference - cell)
Decimation	dividing out a (constant) base number until data point is close to 1	=LOG(cell, constant)
Smoothing	compute (limited) moving average	=AVERAGE(cell:cell)

---

---

# Data (Dis)honesty

---

---

# Activity: Make your own data

[https://apod.nasa.gov/apod/random\\_apod.html](https://apod.nasa.gov/apod/random_apod.html)

- refresh page until you find image you like
- post the date of the image in Discord, first come first serve!
- download image
- drag and drop into: <https://geogebra.org/classic>
- mark the "important" points
  - A > point on object
- Save construction protocol
  - top right settings > download as > construction protocol
- Save image of plot

# Images as data are a problem in research

<https://www.biorxiv.org/content/10.1101/049452v1.full>

- What transformations were done to the images in figure 3
- what scales are used (how was the data transformed) in figure 7?