

# (הרבה) נתונים מספרים סיפורי אגדה: מדע אזרחי כחלון לאוריינות נתונים

פרופ' דני בן-צבי, אוניברסיטת חיפה  
TCSS, 31/7/19



Taking Citizen Science to School  
מרכז לקידום מדע אזרחי בבית הספר



Technologies in  
Education Program



CONNECTIONS  
Developing Students' Statistical Reasoning

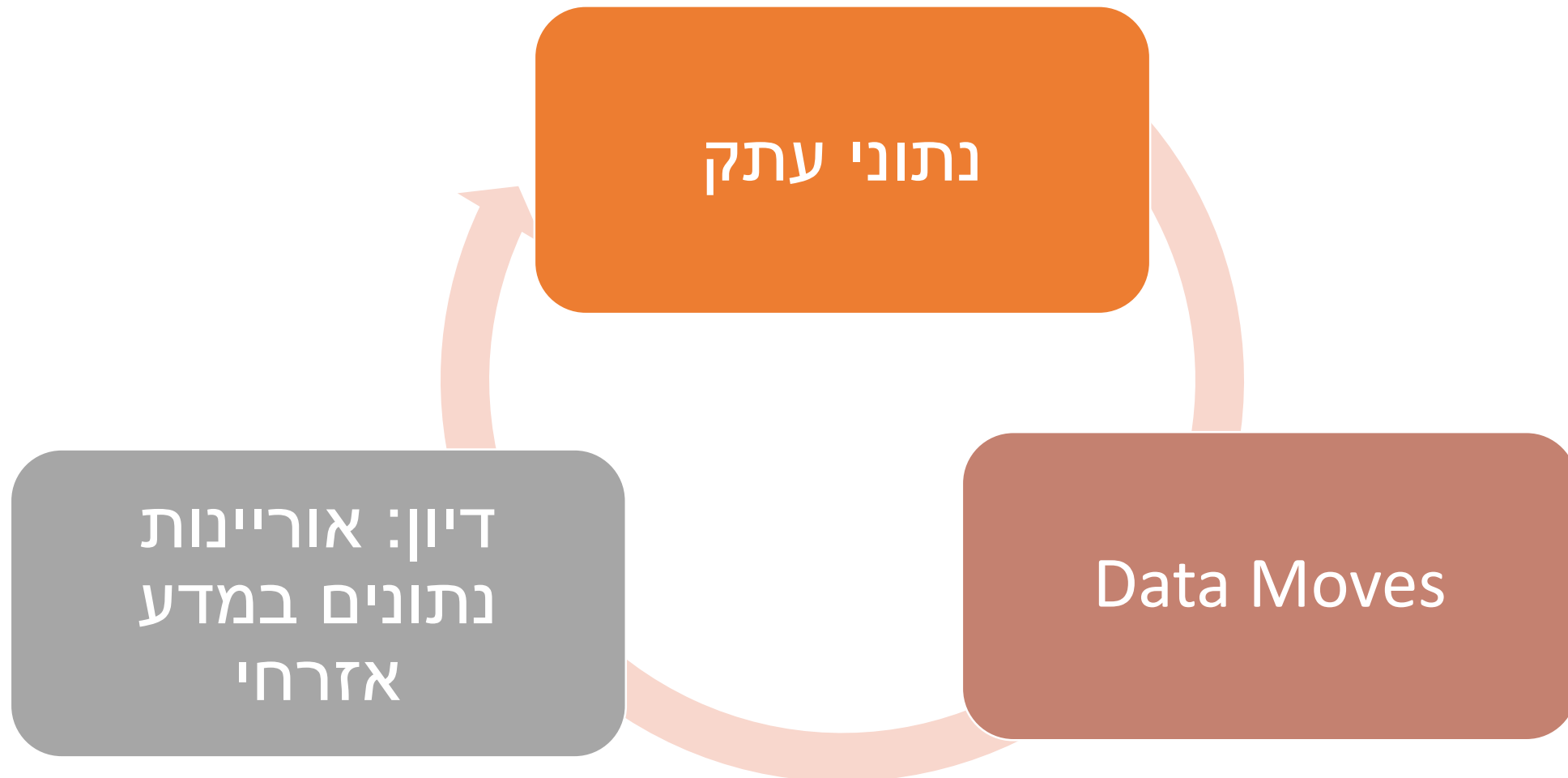


Cool  
Collaboration  
on Learning  
Communities

# מיקוד והנעה



# מתאר ההרצאה





# 1. נתוני עתקו Awash in Data

Gapminder

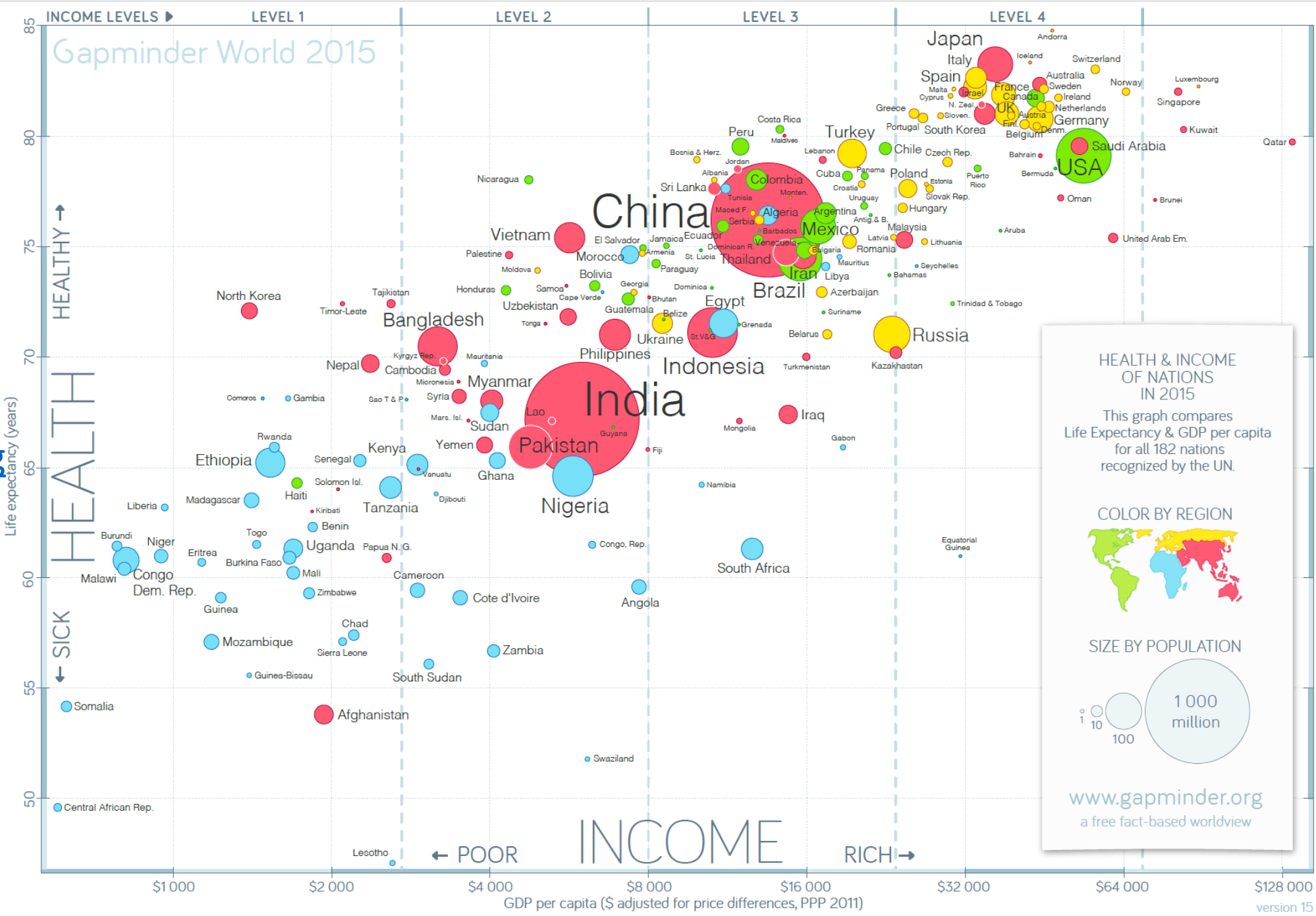
דוגמה 1

# Gapminder

**שאלה: (בזוגות)**  
**חזו וציירו:**  
**מה הקשר בין הכנסה**  
**ובין בריאות?**  
**היום ולפני 100 שנה?**

[www.gapminder.org](http://www.gapminder.org)

**תשובה קצרה:**  
**אנשים עשירים חיים**  
**יותר.**



version 15  
 DATA SOURCES—INCOME: World Bank's GDP per capita, PPP (2011 international \$). Income of Syria & Cuba are Gapminder estimates. X-axis uses log-scale to make a doubling income show same distance on all levels. POPULATION: Data from UN Population Division. LIFE EXPECTANCY: IHME GBD-2015, as of Oct 2016.  
 ANIMATING GRAPH: Go to [www.gapminder.org/tools](http://www.gapminder.org/tools) to see how this graph changed historically and compare 500 other indicators. LICENSE: Our charts are freely available under Creative Commons Attribution License. Please copy, share, modify, integrate and even sell them, as long as you mention: "Based on a free chart from www.gapminder.org".

An aerial photograph of the ocean's surface, showing a dense pattern of white, frothy waves and foam against a deep blue background. The perspective is from directly above, looking down at the water's texture.

Census at School

דאגמא II

# Census at School

<https://ww2.amstat.org/censusatschool/>

- השנה: 2019
- מספר הרשומות שהתקבלו: 75,930

## השאלה:

מה הקשר בין מוטת הידיים והגובה?  
שוחחו עם בן/בת זוג ערכו תחזית  
וציירו.



## U.S. Census at School Questionnaire

The first thirteen questions marked with an asterisk are international questions common to the countries participating in the Census at School program.

1. Are you male or female?\*

Male

Female

2. How old are you?\*

Age (years) \_\_\_\_\_

3. Are you right-handed, left-handed or ambidextrous? (An ambidextrous person is able to use his or her right and left hands equally well.)\*

Right-handed

Left-handed

Ambidextrous

How tall are you without your shoes on?

5. What is the length of your right foot (without your shoe on)? Answer to the nearest centimeter.\*

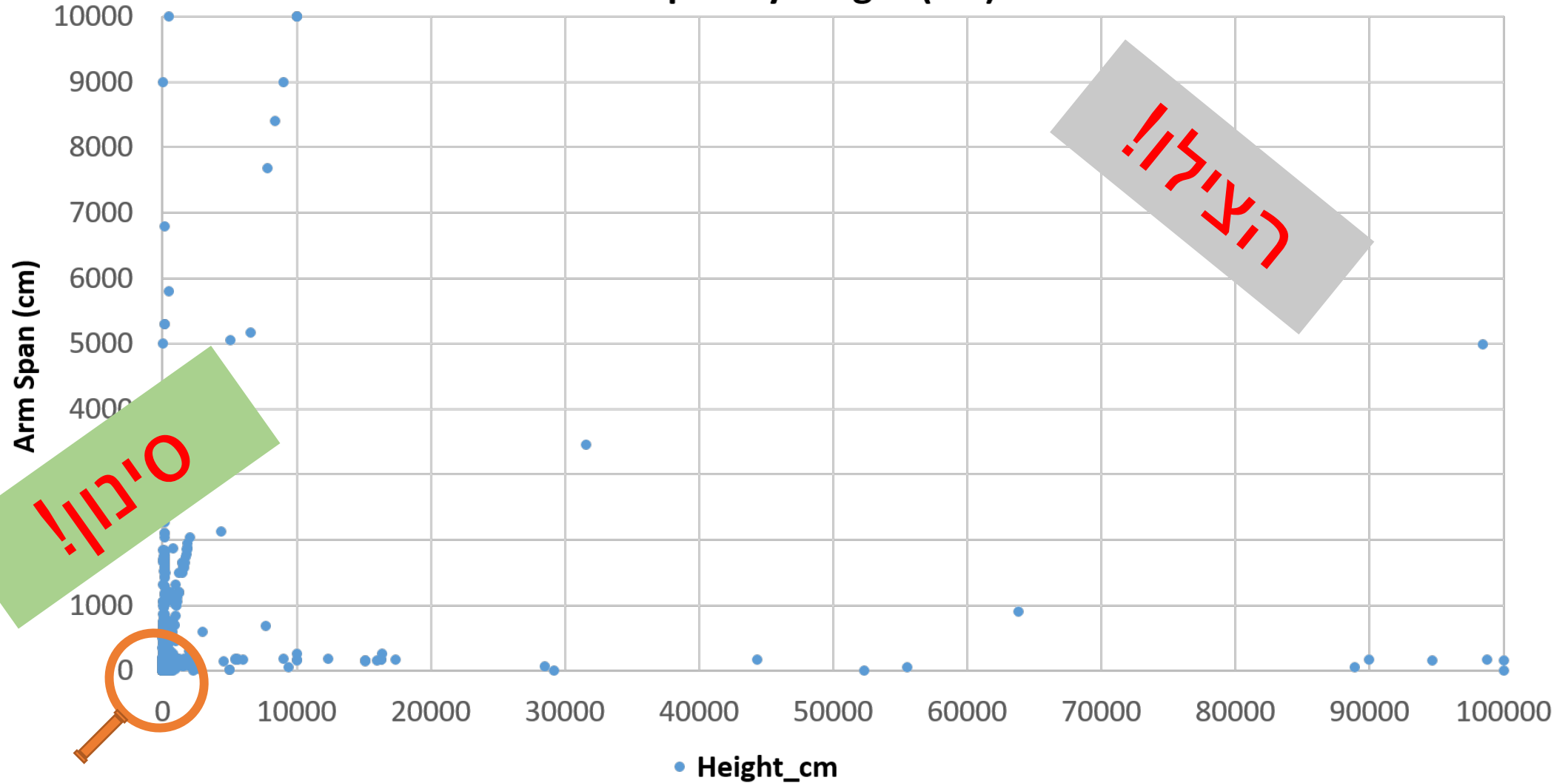
Right foot length (cm) \_\_\_\_\_

What is your arm span?



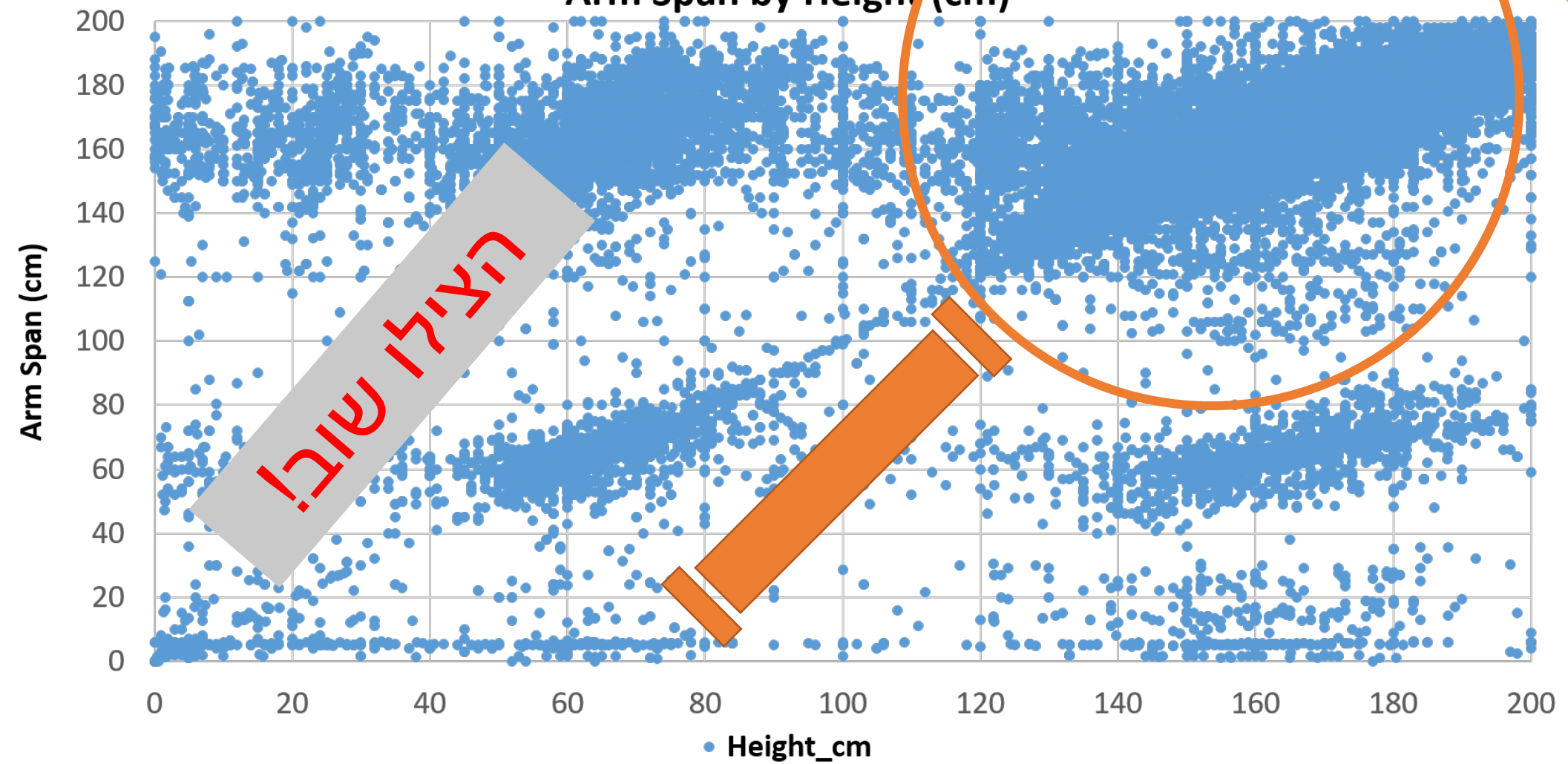
# Census at School

## Arm Span by Height (cm)



# Census at School

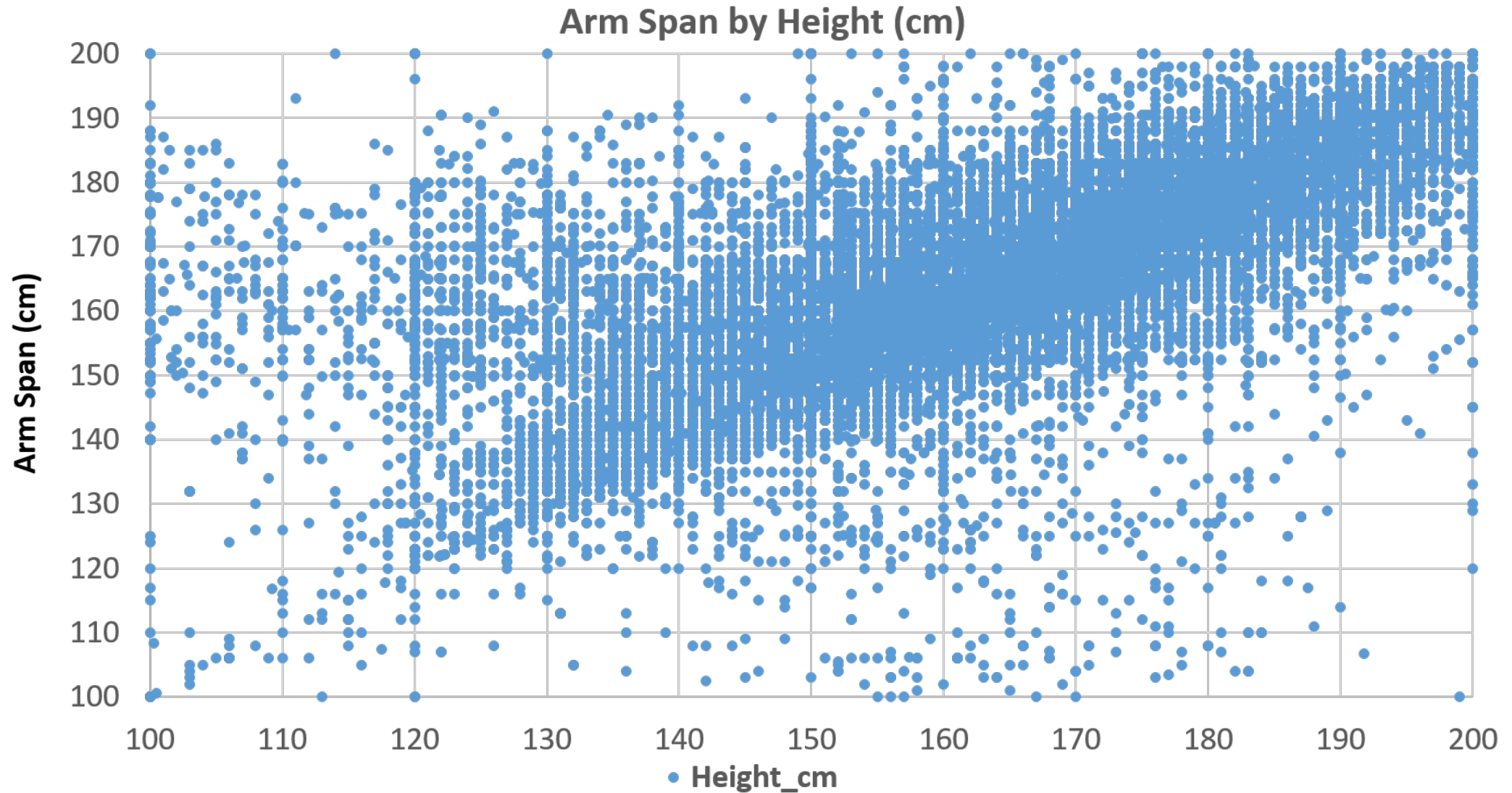
## Arm Span by Height (cm)



הצילו שוב!

א'תלל!

# Census at School



# Census at School: Help to clean the data!

## ניקוי נתונים

עבדו בזוגות

והצדיקו בחירותיכם:

1. אלו רשומות תקינות?
  2. אילו ניתן לתקן וכיצד?
  3. על אלו נאלץ לוותר?
- לכל אחת מהשאלות  
מצאו רשומה אחת בלבד.

#	Response ID	Age (years)	Gender	Arm span (cm)	Height (cm)
1	613090	12	Male	143	153
2	652798	17	Male	1888	188
3	652799	17	Male	1888	188
4	629843	13	Male	88	177
5	651122	17	Female	1.33	153
6	600428	16	Female		
7	600432	15	Male		171
8	638183	10	Male	1.42	1.6
9	660285	17	Female	153	167.6
10	601149	17	Male	92	185
11	613121	99999	Female	99999	99999
12	676958	17	Female	165	163
13	602302	18	Female	2	7
14	623448	14	Female	131	153
15	641603	18	Male	191	184
16	602854	2011	Male	15434	91380
17	654546	11	Female	153	165
18	666160	12	Female	2.9	4.3
19	640912	13	Male	214	245

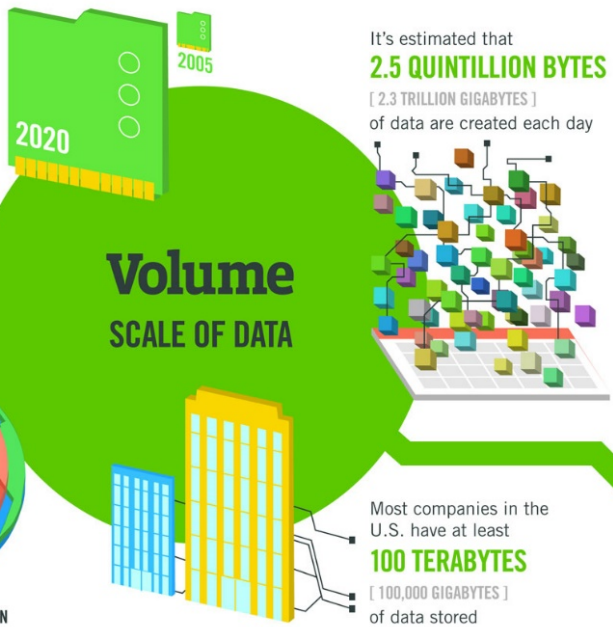
# להיות מוצף בנתונים (Awash in Data)

מהי התחושה של "להיות מוצף בנתונים"?

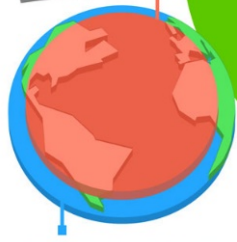
- קל להיות מבולבל על ידי הנתונים ולא להבין את משמעותם
- הכיוון הטוב ביותר אינו ברור והדרך אינה סלולה
- המשימה היא למצוא סדר בכאוס, מחט בערמת שחת, אות ברעש
- הסיטואציה אינה סטנדרטית ולכן היא מזמינה יצירתיות ותחכום

## 40 ZETTABYTES

[ 43 TRILLION GIGABYTES ]  
of data will be created by 2020, an increase of 300 times from 2005



**6 BILLION PEOPLE** have cell phones



WORLD POPULATION: 7 BILLION

The New York Stock Exchange captures

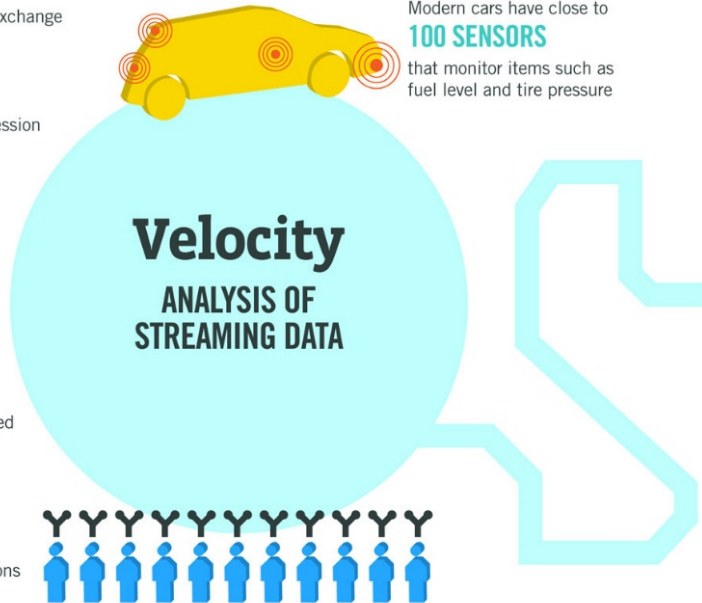
**1 TB OF TRADE INFORMATION** during each trading session



By 2016, it is projected there will be

**18.9 BILLION NETWORK CONNECTIONS**

— almost 2.5 connections per person on earth



# The FOUR V's of Big Data

From traffic patterns and music downloads to web history and medical records, data is recorded, stored, and analyzed to enable the technology and services that the world relies on every day. But what exactly is big data, and how can these massive amounts of data be used?

As a leader in the sector, IBM data scientists break big data into four dimensions: **Volume, Velocity, Variety and Veracity**

Depending on the industry and organization, big data encompasses information from multiple internal and external sources such as transactions, social media, enterprise content, sensors and mobile devices. Companies can leverage data to adapt their products and services to better meet customer needs, optimize operations and infrastructure, and find new sources of revenue.

By 2015 **4.4 MILLION IT JOBS**

will be created globally to support big data, with 1.9 million in the United States



As of 2011, the global size of data in healthcare was estimated to be

**1.5 QUINILLION BYTES** [ 1.3 TRILLION GIGABYTES ]

By 2014, it's anticipated there will be

**420 MILLION WEARABLE, WIRELESS HEALTH MONITORS**

**4 BILLION+ HOURS OF VIDEO**

are watched on YouTube each month



**30 BILLION PIECES OF CONTENT**



**Designed vs. Organic Data**

**1 IN 3 BUSINESS**

information is accurate



data quality costs the US

around **\$100 BILLION A YEAR**

**Data vs. Information**

**27% OF RESPONDENTS**

in one survey were unsure of how much of their data was inaccurate

**information-to-data ratio**



# 2. Data Moves לנסות לעשות סדר בכאוס

# פעולות לשינוי נתונים (data moves)

- כאשר טירונים מתמודדים עם נתוני עתק (big data) אורגניים או מעוצבים, יש לעתים קרובות צורך במניפולציה של הנתונים (data wrangling, data moves) ולהצדיק אותה - כדי שאפשר יהיה להשיב על שאלות מסקרנות, למשל:
- לסנן (filter) את הנתונים, למשל, כי הם כוללים מידע שגוי או לא רלוונטי למטרות המחקר,
- למזג (merge) קובץ נתונים עם קובץ נתונים אחר, או
- לקבץ (group) את הנתונים לקבוצות חדשות, או ליצור מדדים (measure) חדשים.
- למרות שפעולות אלו שכיחות, הן לא נלמדות כחלק חיוני של הוראת נתונים (Erickson et al., 2019; Kandel, et al. 2011).



# פעולות לשינוי נתוני עתק (data moves)

- ✓ Cleaning (ניקוי)
- ✓ Filtering (סינון)
- ✓ Grouping (קבוצות)
- ✓ Summarizing (סיכום)
- ✓ Calculating (חישוב)
- Merging (מיזוג)
- ✓ Making Hierarchy (יצירת מדרג)

# מה המטרות של שינויי נתונים (Data moves)?

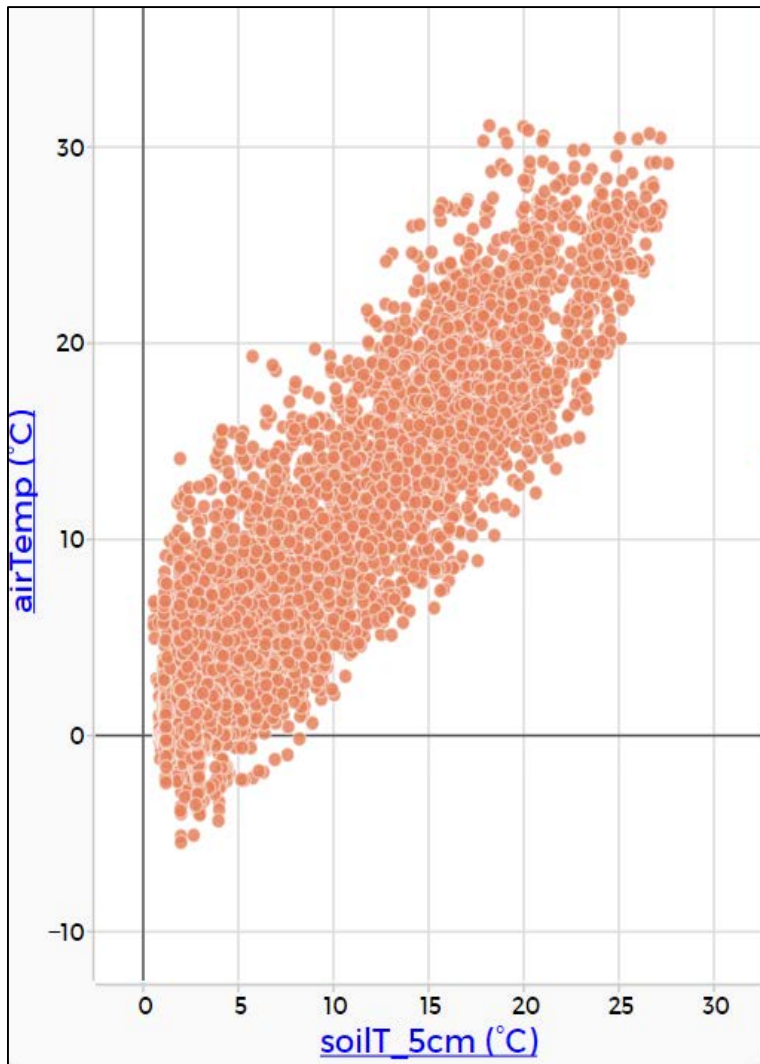
- לקדם את ההבנה של הנתונים והחקירה
- להכין או לפשט את מערך הנתונים (dataset)
- להסיר רשומות שגויות, לא רלוונטיות
- לארגן מחדש את מערך הנתונים
- להגדיל את המורכבות של מערך הנתונים כדי לשקף את מטרות המחקר

# סינון (Filtering)

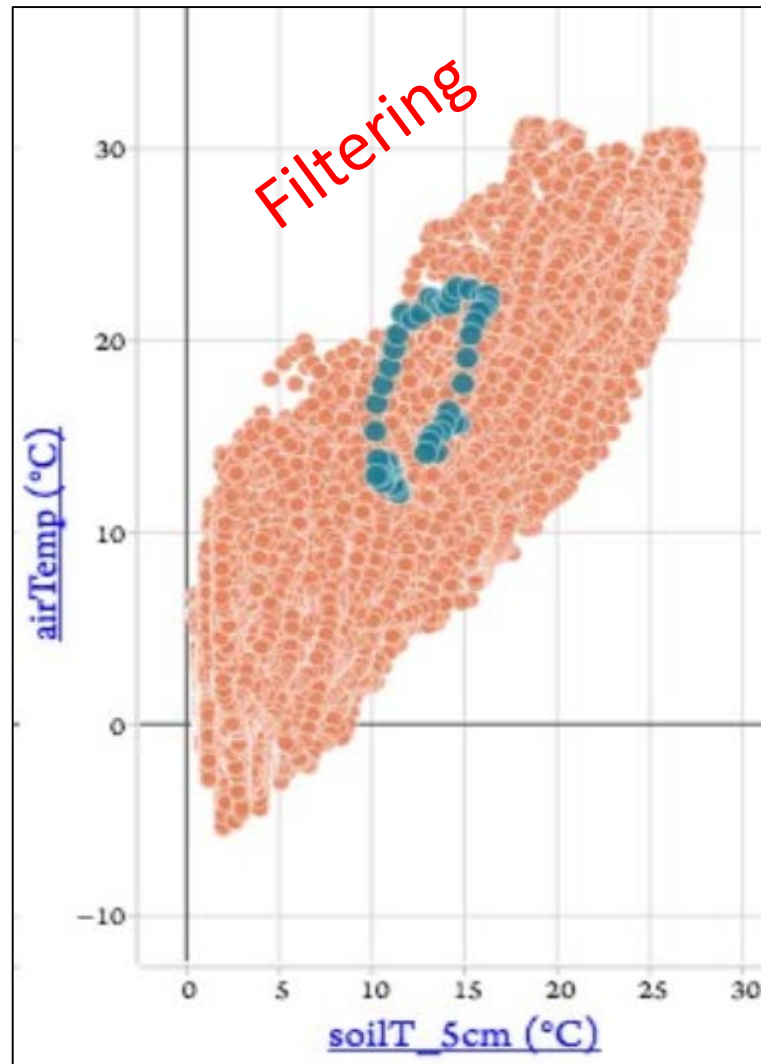


- נתוני עתק על יער Blodgett בקליפורניה
- נאספו על ידי California Forest Research Station (Thomson, 2018)
- כל רשומה (case) מכילה תכונות רבות, ביניהן **טמפרטורת האוויר** ו**טמפרטורת האדמה** (בעומק 5 ס"מ), שנמדדו באותו זמן
- התכונות הללו נאספו כל 30 דקות בשנת 2000 – 17,500 רשומות בסה"כ.

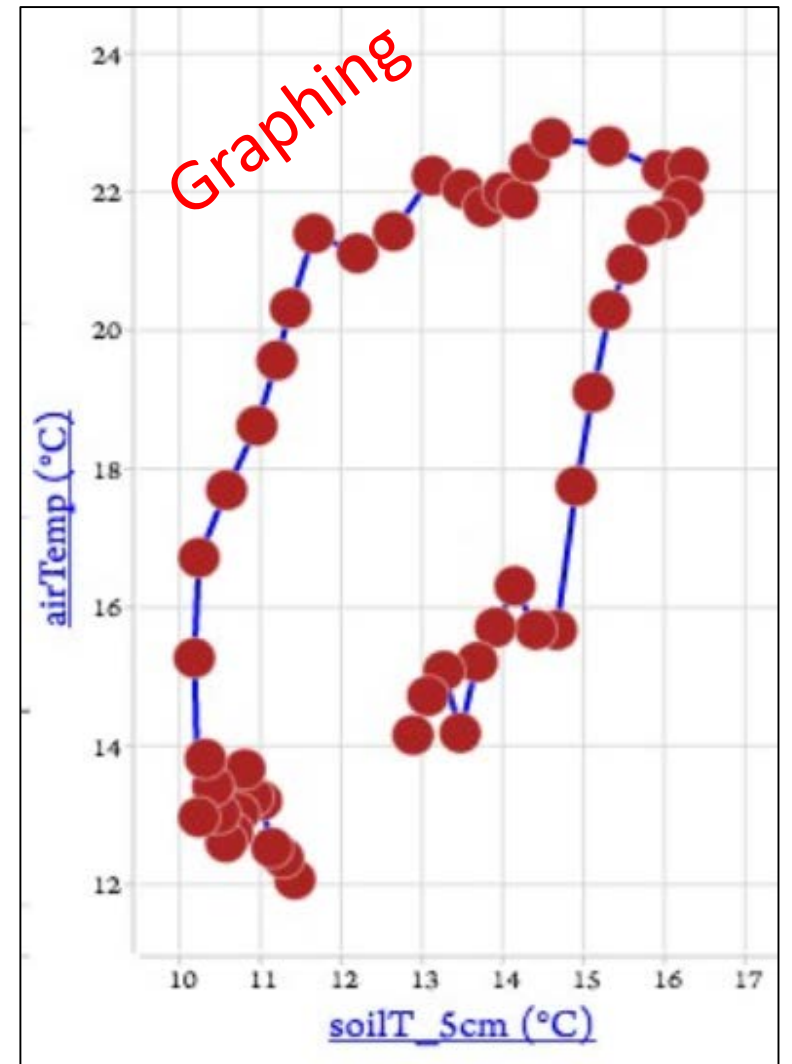
**המטרה: הקשר בין הטמפרטורה באוויר ובין טמפרטורת הקרקע  
חזו וציירו בזוגות.**



Air temperature versus soil temperature (5 cm depth) every 30 minutes for the year 2000.



May 19 2000 data highlighted.



A graph showing only the data from May 19, 2000, rescaled to fill the plot.

# סקר בריאות ותזונה של ילדים

- סקר מתמשך של National Center for Health Statistics (NCHS) על בריאות ותזונה של אזרחי ארה"ב המשלב ראיונות עם בדיקות פיזיות
- הנתונים: 800 ילדים בגילאים 5-19 בשנת 2013

## מטרות

- קשר בין גובה ומגדר
- מגמות גדילה של ילדים

National Health and Nutrition Examination Survey ("[NHANES](#)"; CDC 2003)

# מדגם הנתונים הגולמיים על בריאות ותזונה

nhanes ages 5 to 19							
Cases (800 cases)							
index	Gender	Age	Weight	Height	BMI	UpperArm	Pulse
1	Male	18	66.8	176	21.57	38.1	54
2	Female	17	58.7	164.7	21.64	35.5	50
3	Female	5	19.8	114.9	15	24.9	
4	Male	10	43	141.3	21.54	32.2	72
5	Female	13	48.4	155.4	20.04	34.4	64
6	Female	12	52.1	150.8	22.91	33.4	94
7	Female	10	55	150.2	24.38	31.5	92
8	Male	19	80.4	187.8	22.8	39.3	52
9	Female	14	54.2	160	21.17	33.4	106
10	Female	8	22	122.1	14.76	25	102

Case  
רשומה

Attribute  
תכונה

Value  
ערך



# קיבוץ (Grouping)

המטרה:

ללמוד על גובה על פי מגדר

<i>Genders (2 cases)</i>			<i>Cases (800)</i>			
index	Gender		index	Age	Weight	Height
1	Male	+	411 Cases			
2	Fema...	+	389 Cases			

Grouping




# קיבוץ וחישוב (Grouping and calculating)

המטרה:

ללמוד על גובה על פי מגדר

האמצעי:

1. יצירת 2 קבוצות מגדר
2. חישוב גובה ממוצע לכל קבוצה
3. השוואה בין הממוצעים

Genders (2 cases) 		
index	Gender	MeanHT
1	Female	149.59
2	Male	159.53

Grouping

Calculating

Genders (30 cases)			
index	Age	Gender	meanHT
1	5	Female	112.15
2	5	Male	113.85
3	6	Male	119.3
4	6	Female	119.28
5	7	Male	129.71
6	7	Female	127.89
7	8	Female	131.83
8	8	Male	133.32
9	9	Female	136.56
10	9	Male	137.17
11	10	Male	143.17
12	10	Female	146.83
13	11	Male	151.2
14	11	Female	153.45
15	12	Female	157.4
16	12	Male	156.11
17	13	Female	157.74

Grouping \* 2  
Calculating

# קיבוץ וחישוב

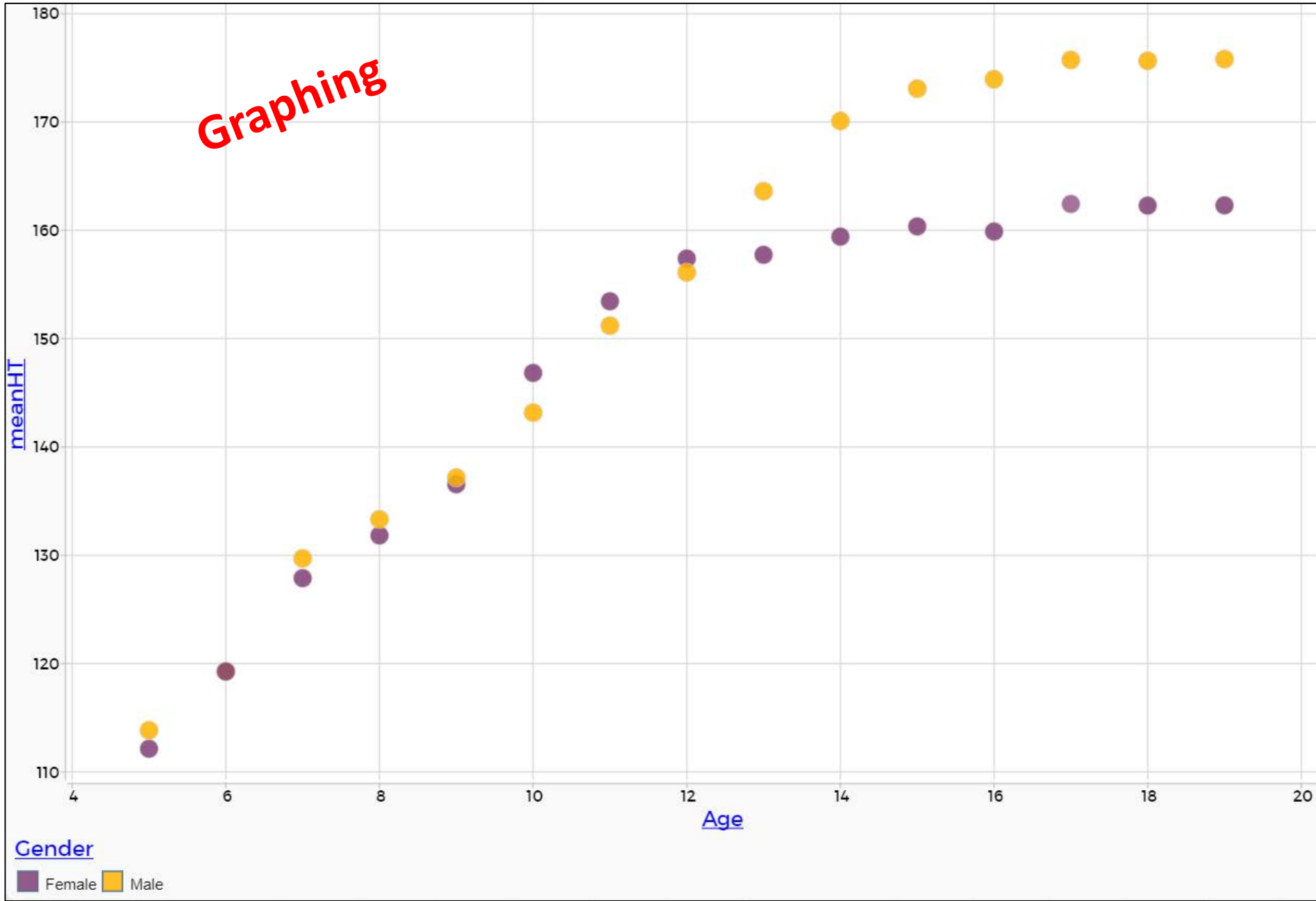
המטרה:

מגמות הצמיחה לגובה של ילדים

האמצעי:

1. יצירת 30 קבוצות של גיל ומגדר
2. חישוב הגובה הממוצע של כל קבוצה

Graphing



# Graphing

המטרה:

מגמות הצמיחה  
לגובה של ילדים

# פעולות לשינוי נתוני עתק (data moves)

- ✓ Cleaning (ניקוי)
- ✓ Filtering (סינון)
- ✓ Grouping (קבוצות)
- ✓ Summarizing (סיכום)
- ✓ Calculating (חישוב)
- Merging (מיזוג)
- ✓ Making Hierarchy (יצירת מדרג)

# פעולות לשינוי נתוני עתק (data moves)

## שינוי תוכן

- פעולה שמשנה את התוכן, המבנה והערכים של מערך נתונים (למשל, סינון)

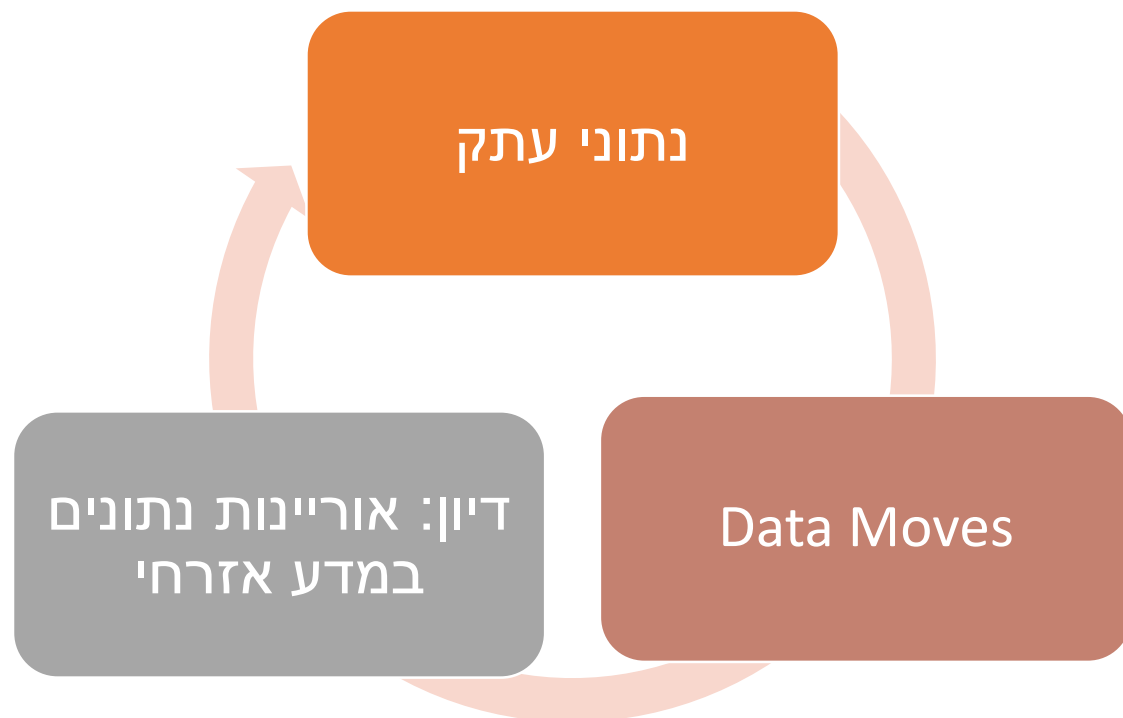
## שינוי מבנה

- פעולה שמשנה את היחסים בין רשומות, תכונות וערכים (למשל, קיבוץ, הירארכיה)

## שינוי ערכים

- פעולה שמשנה את הערכים בתאים (למשל, חישוב, ניקוי)

# דיון



- מה כל זה אומר לי לגבי פרויקט המדע האזרחי שלי?
- אילו צעדים פדגוגיים עלינו לנקוט כדי לטפח אוריינות נתונים במסגרת מדע אזרחי
- הערות ושאלות אחרות



Taking Citizen Science to School  
"מרכז לקידום מדע אזרחי בבית הספר"



CONNECTIONS  
Developing Students' Statistical Reasoning



Cool  
Collaboration  
on Learning  
Communities

# מקווה שלא טבעתם באוקיינוס הנתונים...

## תודה!

[dbenzvi@univ.haifa.ac.il](mailto:dbenzvi@univ.haifa.ac.il), [dbz.edtech.haifa.ac.il](mailto:dbz.edtech.haifa.ac.il)