

bit.ly/teesdss2018

Data Moves

—

Tim Erickson



This is about education.

Perspective: starting to
learn DS without coding.

See [Rebecca Nugent](#)

The team uses **CODAP**

Common **O**nline **D**ata **A**nalysis **P**latform
—gesture-based, free, runs in the
browser, open source, open architecture,
JavaScript, NSF, etc. codap.concord.org

What are the building
blocks — the atoms —
of data science?

Looking for tasks that
“smell” like data science

We noticed...

- (1) We took particular actions in order to accomplish a variety of tasks.
- (2) The actions were computery.
- (3) They never appeared in stats class.

We called them *data moves*.

Useful construct? Let's see.

Aha!

Unacknowledged prerequisites?

Let's see some.

<https://codap.concord.org/releases/latest/static/dg/en/cert/index.html#shared=36675>

—

First three data moves:

- (1) Filter or subset
- (2) Group
- (3) Aggregate (aka Summarize)

These are all interrelated...

...and students are not fluent in what these mean, conceptually.

Three more data moves:

(4) Calculate (new attribute)

(5) Merge or Join

(6) Make hierarchy

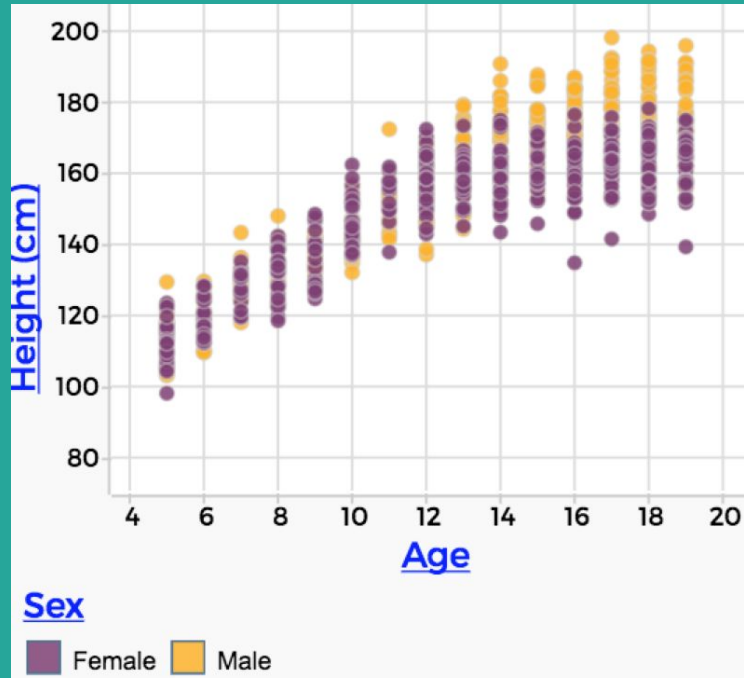
Let's see hierarchy.

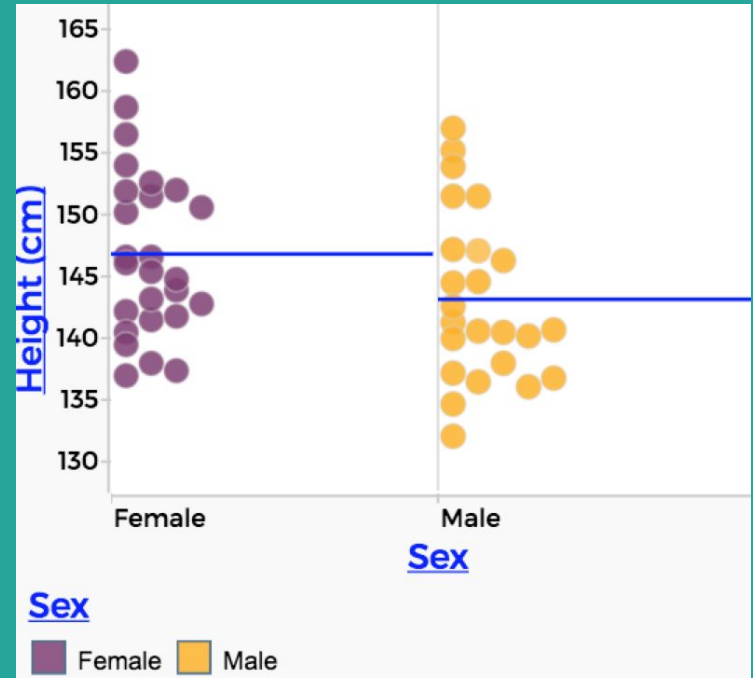
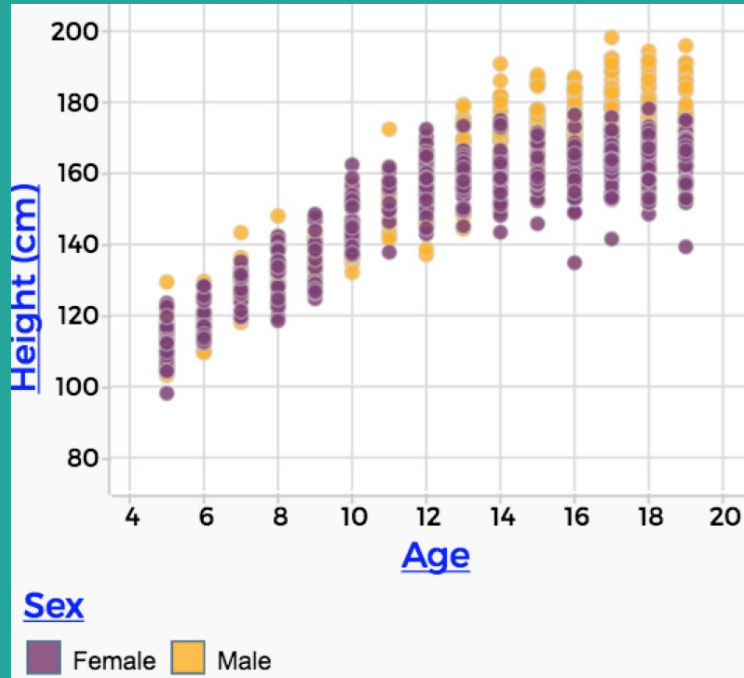
<https://codap.concord.org/releases/latest/static/dg/en/cert/index.html#shared=36675>

—

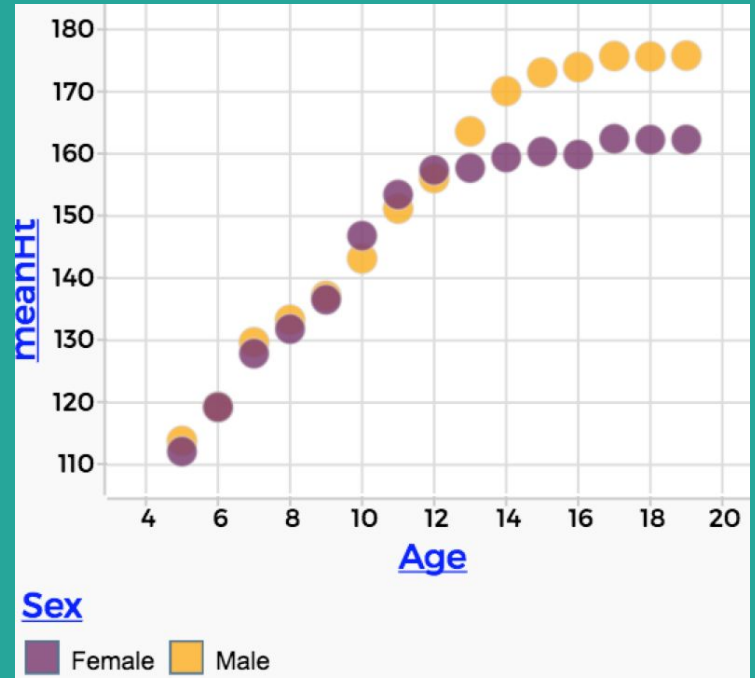
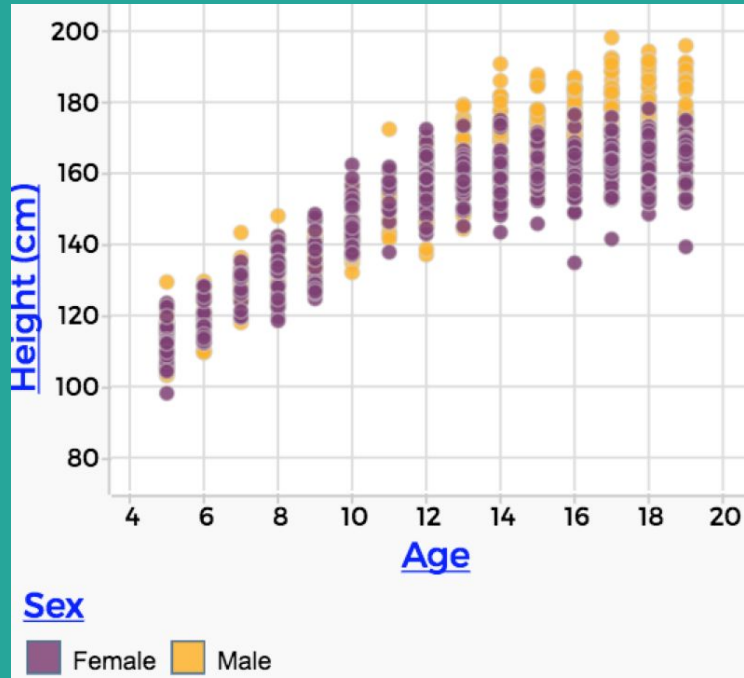
About hierarchy:

- Powerful! Surprising!
- Aggregate measures are [just] new columns at a higher level
- It's a form of iteration
- We keep finding new applications, e.g., recoding

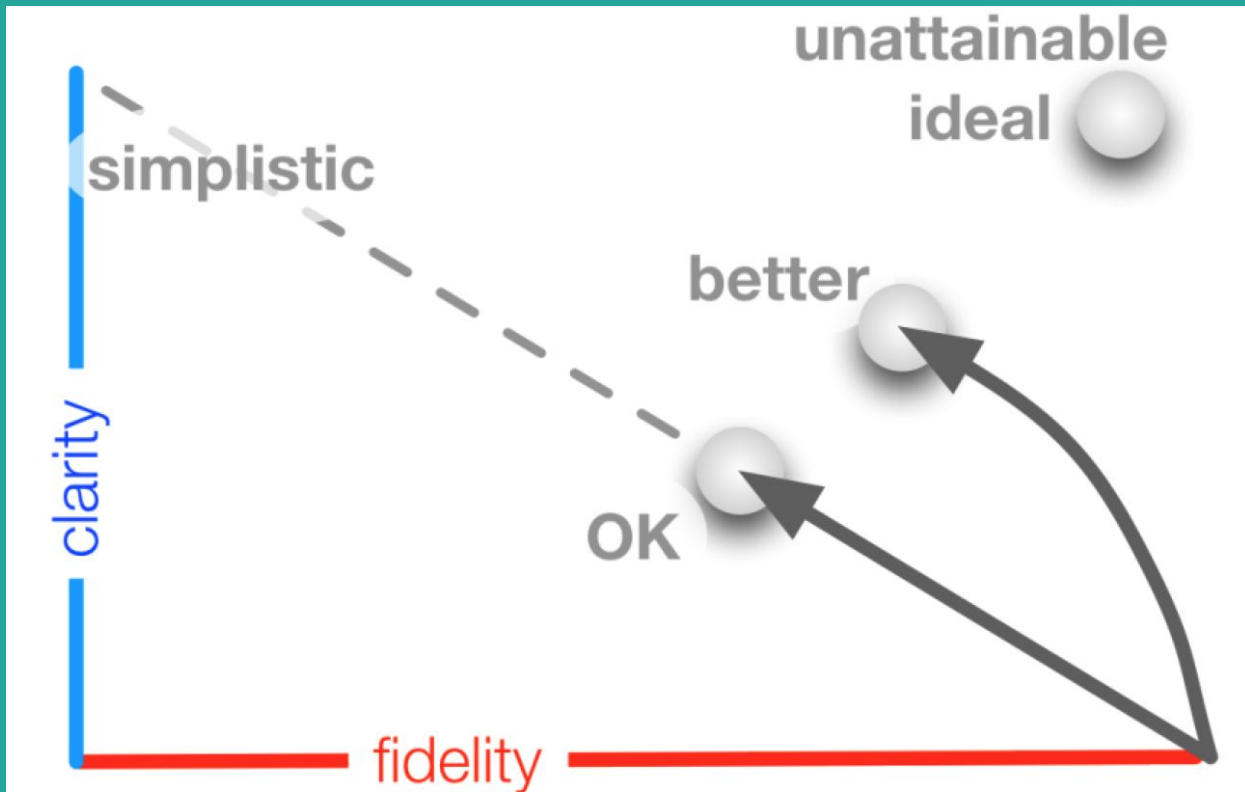




Reflection on **filters** and aggregation.



Reflection on filters and **aggregation**.



We struggle to balance clarity and fidelity.

Other candidates include:

- (7) Make convenience attribute
- (8) Sort
- (9) Stack or pivot
- (10) Munging and cleaning actions

And then we learned about
the tidyverse...

...and behold, our principal data moves
turned out to be tidyverse verbs.

Recommendations:

- Think about what data moves you are asking students to use
- Consider naming them explicitly
- Use them as part of a transition to coding and reproducibility

Your Tax Dollars at Work

This material is based upon work supported by the National Science Foundation under Grant No. IIS-1530578. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. But we can always hope.

