

# Tableau for Academic Researchers

Have you ever wanted to explore your data without writing hundreds of lines of code in R, Python, SAS, Matlab, etc.? (I have, and I love coding!)

Let's explore Tableau as an alternative tool for academic researchers who want to create beautiful and comprehensible visualizations of their data.



*Dan Turner*

*Data Science Consultant*

*Ph.D Candidate, Dept. of Linguistics*

*Northwestern University*

*Research Computing Services, NUIT*

*Summer 2022 Workshop*

# Background

## About me

(Dan Turner)

[dturner@u.northwestern.edu](mailto:dturner@u.northwestern.edu)

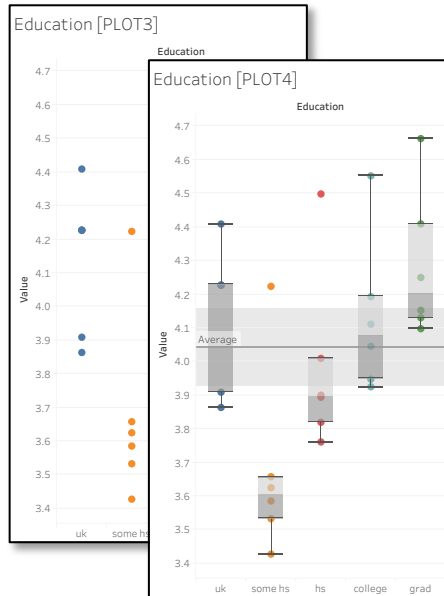
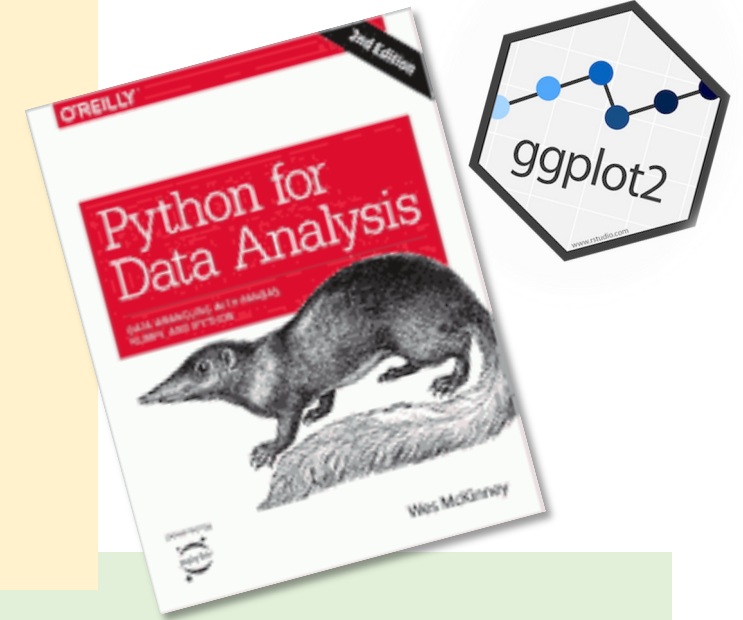
- 5<sup>th</sup> year Linguistics Ph.D Candidate with some Natural Language Processing (NLP) in my background
- My coding languages in order:
  1. Python
  2. Haskell → → →
  3. R
  4. Javascript



- Haskell is a basic functional programming language that compiles. "Correctness" is required.
- I had to learn it in order to TA a class, but ultimately I did not find it useful day-to-day.
- It inspired me to learn R, which is the language I use (and love) the most.

# Flashback to 2018...

- About 4 years ago, I took a Data Visualization course that required learning basic Tableau, since it was designed for people without coding experience.
- It took some time to get over the feeling that Tableau was more about aesthetics than visualizing robust statistics.



- Now, I think that having a graphical user interface (GUI, “gooey”) can make data exploration intuitive, fast, and yields a more attractive (and hopefully more interpretable) data visualization.
- Tableau also has lots of tools that make it handy outside the context of plots too, such as playing nicely with a large variety of data formats. I’m looking at you, JSON.

# Roadmap

## Part 1: Getting started

- Data types
- Panels & workflow
- Plot types
- Common Aggregations

## Part 2: Next steps

- Analytics [Boxplots, regression, k-means clustering]
- Dashboards
- Exporting figures and data

## Part 3: Case study

- Thoughts about reproducibility
- How I'm using Tableau on a project right now

# Want to try for yourself?

## Before you begin

1. Install Tableau, which is free for students & educators
2. Download the files for the workshop from this link:  
**[github.com/turnerdan/tableau\\_4\\_academic\\_research](https://github.com/turnerdan/tableau_4_academic_research)**
3. Open Tableau and finish setup by entering your license (may not be needed)

## Loading data

- Tableau has its own export format, which is included along with the raw data files in the Github repository
- To open raw data, you need to “Connect” the file as a “Data Source” in Tableau.

# Notes before we start



I'll be watching the chat for questions. Some questions might have to be answered after the session because of time constraints.

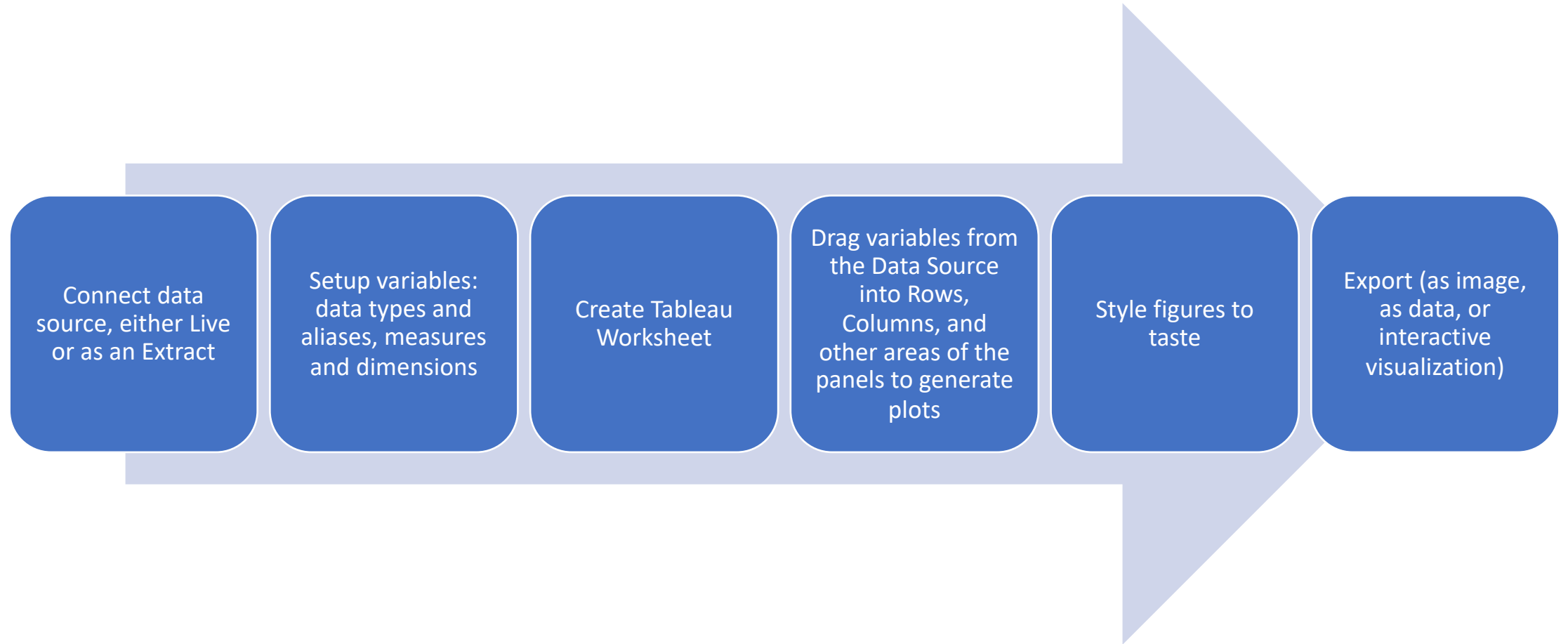


If my internet connection dies mid-workshop, I'll be back as soon as possible. Please give me 10 minutes to fix any issues and return.

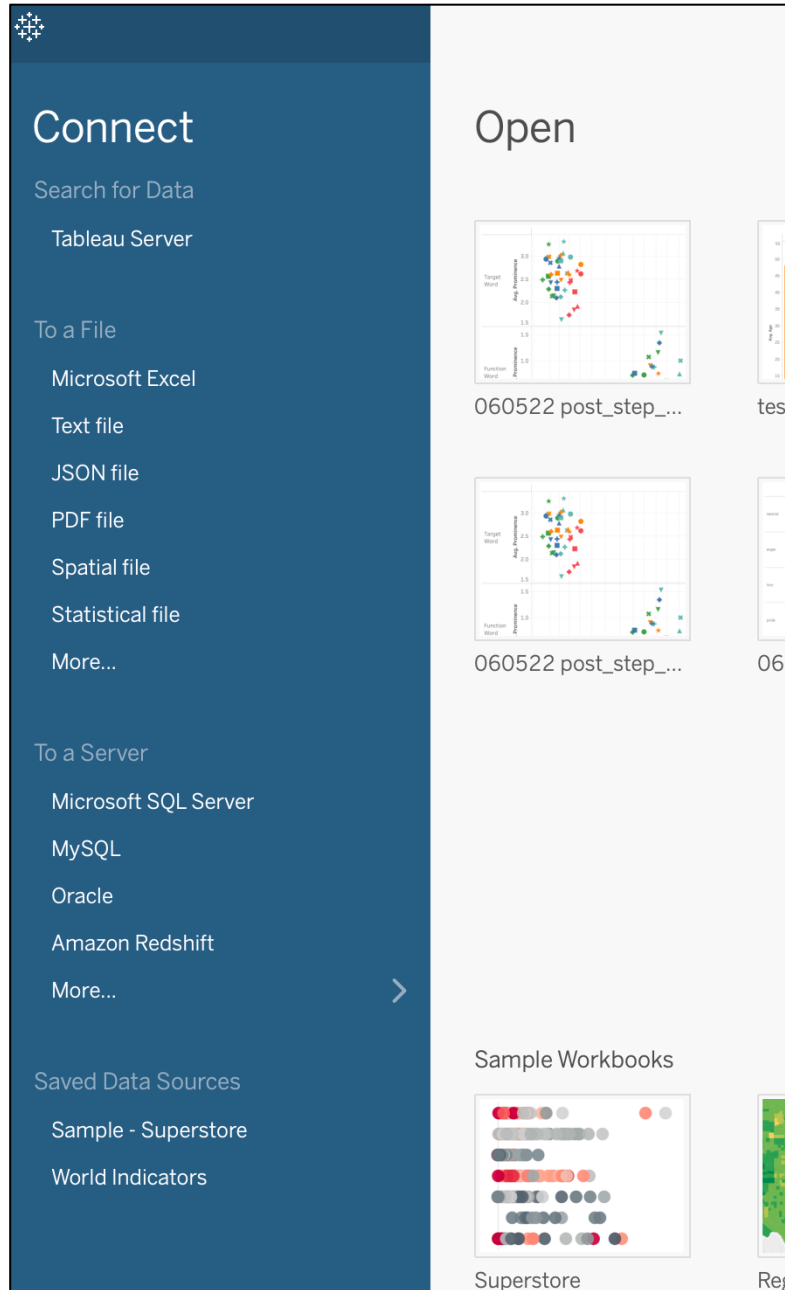
# Part 1: Getting started

General workflow, data types, choosing plot types, common aggregations

# Simplified general workflow







# Some possible data sources & types

- CSV, Excel
- R, SAS, SPSS
- JSON, SQL
- Text files, PDFs
- Spatial/geospatial

Connection types: Live versus Extract

# About Dataset #1

- The data is the results of a quiz that measures “nature relatedness”, which has to do with a person’s attitudes toward ecologically sustainable behavior.
- Data collected from an open online form at [openspsychometrics.com](http://openspsychometrics.com)

Nisbet, E. K., & Zelenski, J. M. (2013). The NR-6: A new brief measure of nature relatedness. *Frontiers in Psychology*, 4.

<https://doi.org/10.3389/fpsyg.2013.00813>

Demographics

Age	Country	Education	Gender	Major	Married	Race	Urban	Engnat	Familysize	Hand	Introelapse	Number of Records	Orientation
33	US	hs	Female		Currently	white	rural	1	2	1	5	1	2
28	US	college	Female	Fine art economics, law	Never	white	urban	1	3	1	3	1	2
39	RO	hs	Female		Currently	white	urban	2	4	2	25	1	1
17	US	hs	Other	NA	Never	other	rural	1	11	2	2	1	2

Answers

Q1A1	Q1E	Q1I	Q2A1	Q2E	Q2I	Q3A1	Q3E	Q3I	Q4A1	Q4E	Q4I	Q5A1	Q5E	Q5I	Q6A1	Q6E	Q6I
5	4404	2	1	4019	6	5	2968	5	5	2950	3	5	2648	4	5	5635	1
5	9284	2	3	6270	6	5	4276	4	5	2413	3	5	5880	1	5	2922	5
4	15199	1	4	8577	5	5	7797	2	5	2988	4	5	3588	3	5	4148	6
2	6895	1	2	3825	6	5	4488	5	3	2939	4	4	6159	2	5	2237	3
2	11741	1	4	4340	2	3	9536	4	5	3219	5	2	6543	3	2	3467	6

Personality

Tipi1	Tipi10	Tipi2	Tipi3	Tipi4
5	1	6	6	5
6	1	3	6	3
2	2	6	7	5
1	1	5	6	7
2	3	2	6	5

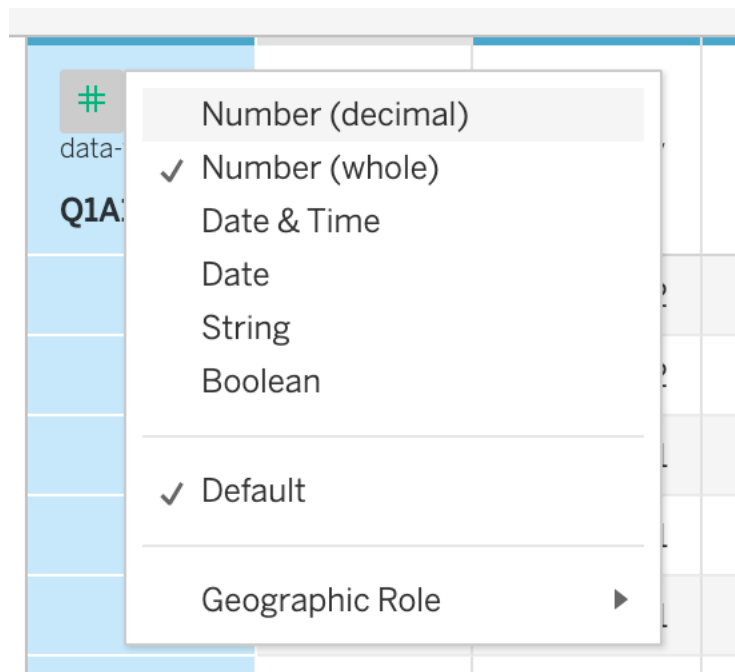
Linguistic knowledge

Vcl2	Vcl3	Vcl4	Vcl5
1	0	1	1
1	0	1	1
1	1	1	1
1	1	1	1
1	0	1	1

# Set data types and variable aliases (labels)

## Data types

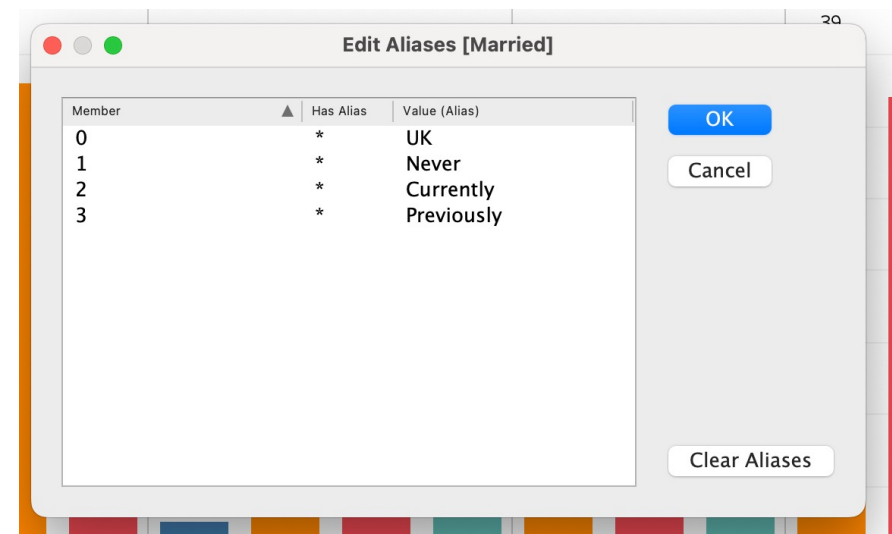
- When you Connect a Data Source, always make sure your data is in the right type.



## Aliases

In the Data Source screen, you can rename variables.

Variables can also have “Aliases” which makes it easy to mask one label with another across Tableau.



# Measures versus Dimensions

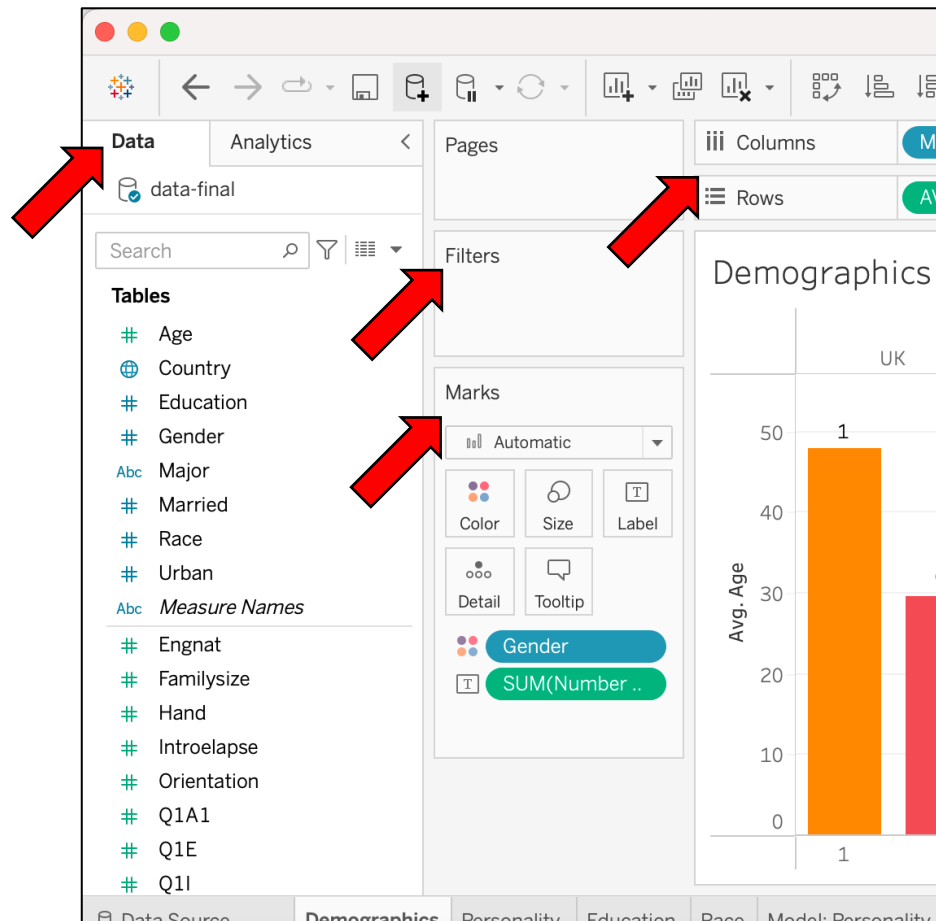
Measures are typically response variables  
(independent variable)

Dimensions are typically predictor variables  
(dependent variable)

- This concept is so important it's always color-coded in the software.
- Tableau doesn't know the difference and setting up the data correctly,
  - However, on import it's assumed that text-type fields and date-type fields are Dimensions and numeric-type fields are Measures.

# Starting a new Tableau Worksheet

Left side



- Data panel
- Filters panel
- Marks panel
- Rows & Columns

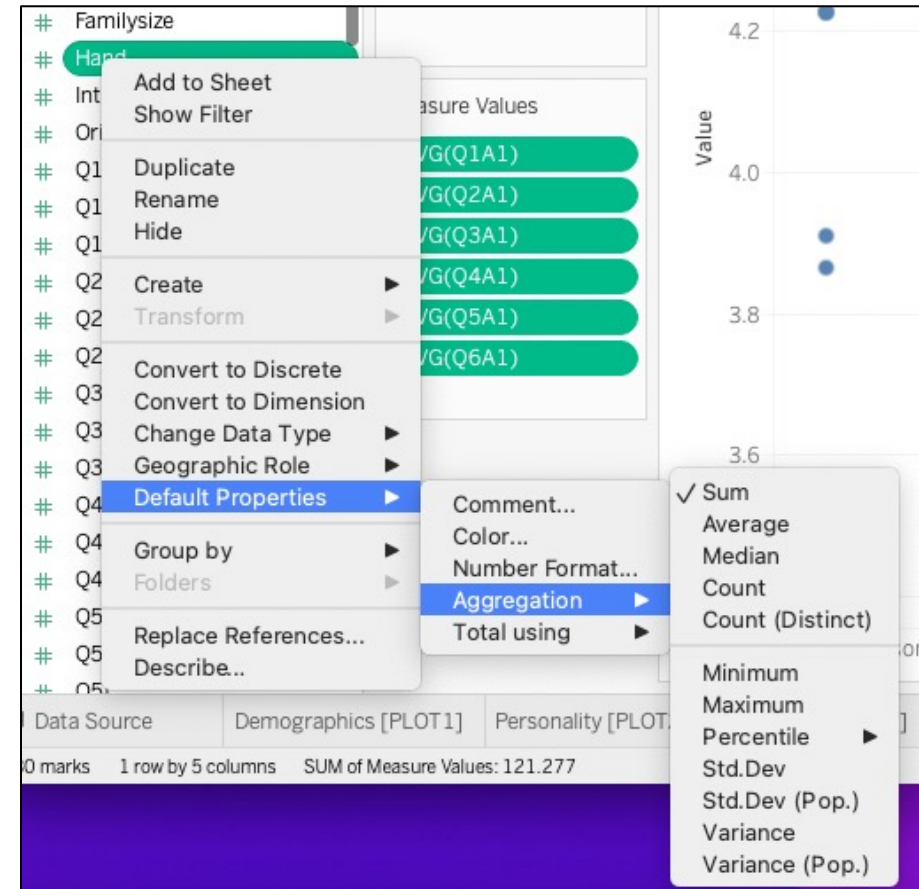
Depending on your needs, you'll need to add Measures and Dimensions to Rows and Columns.

Right side



# Aggregating (grouping & summarizing) data

- Tableau applies default data aggregations to certain data types by default.
  - If your data has two variables, Shoe Size (continuous/numeric) and Hair Color (discrete/text string), Tableau will automatically calculate the Sum of Shoe Size for each Hair Color.
- Changing the default aggregation from Mean to something else is easy, but it's deep in the menus →



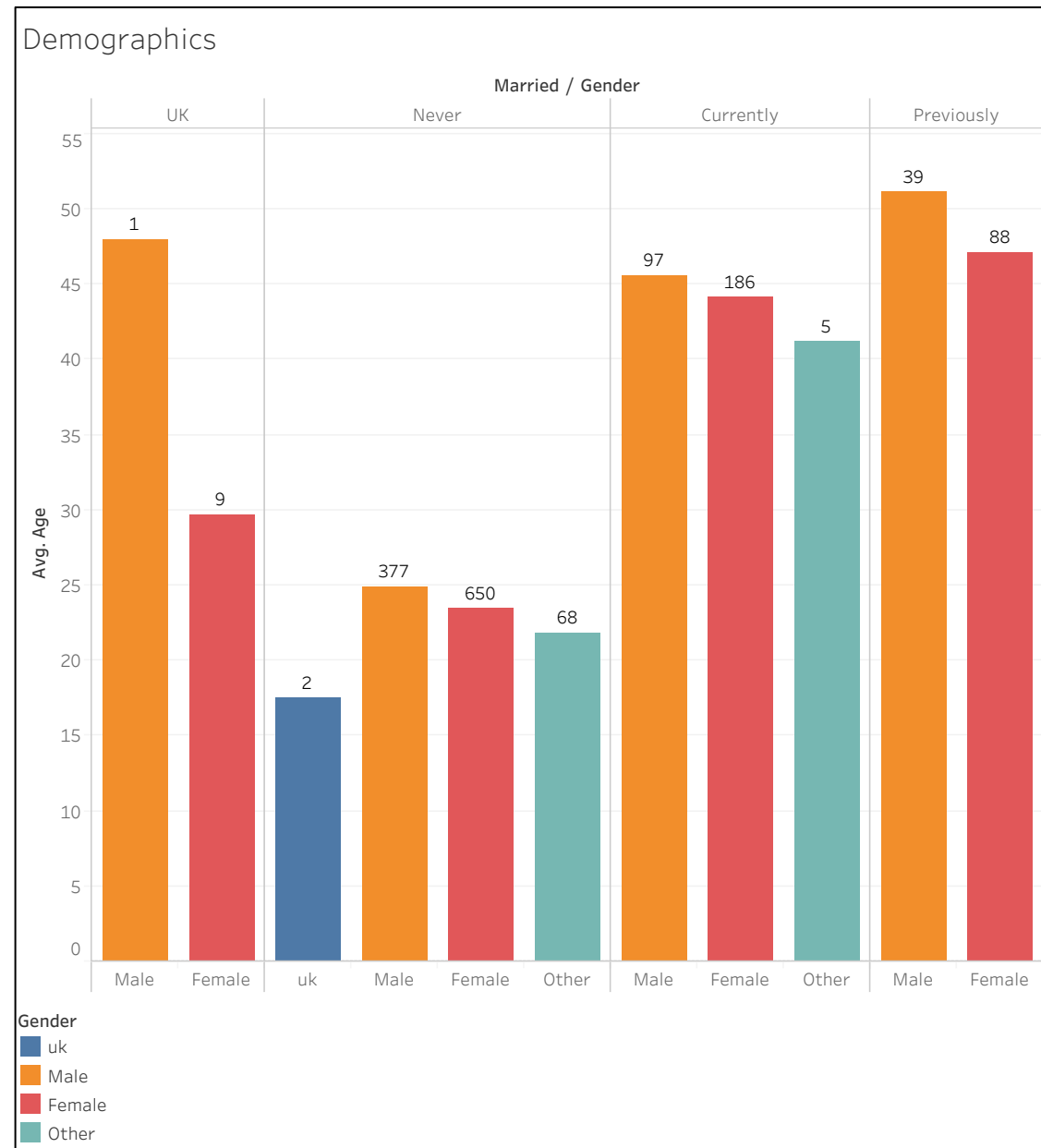
# Plot #1

A bar plot needs 1  
continuous variable...

**Measure, Y axis, Rows**

...and at least 1  
categorical variable...

**Dimension, X axis,  
Columns**

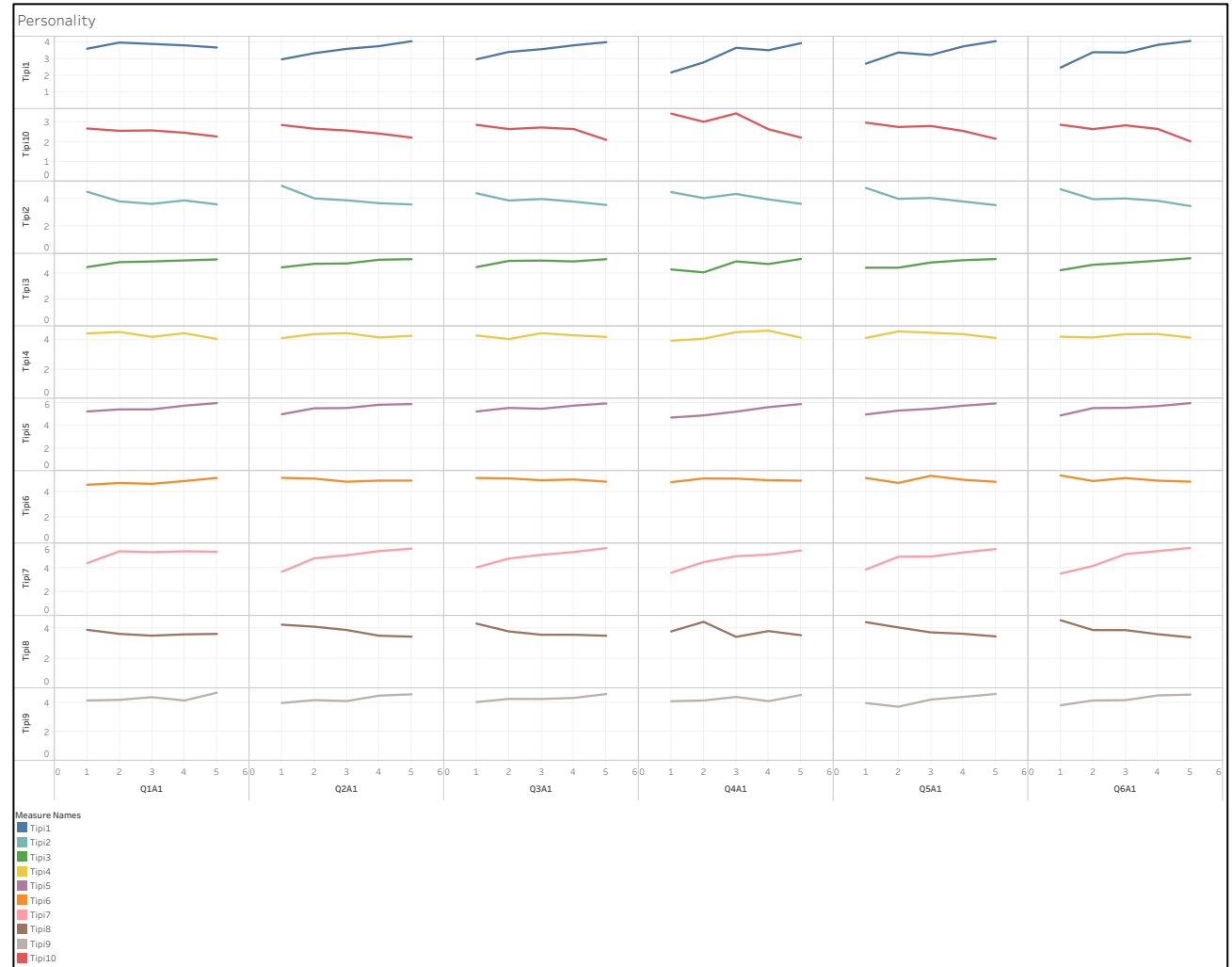




# Plot #2

**Line graphs typically compare 2 or more continuous variables.**

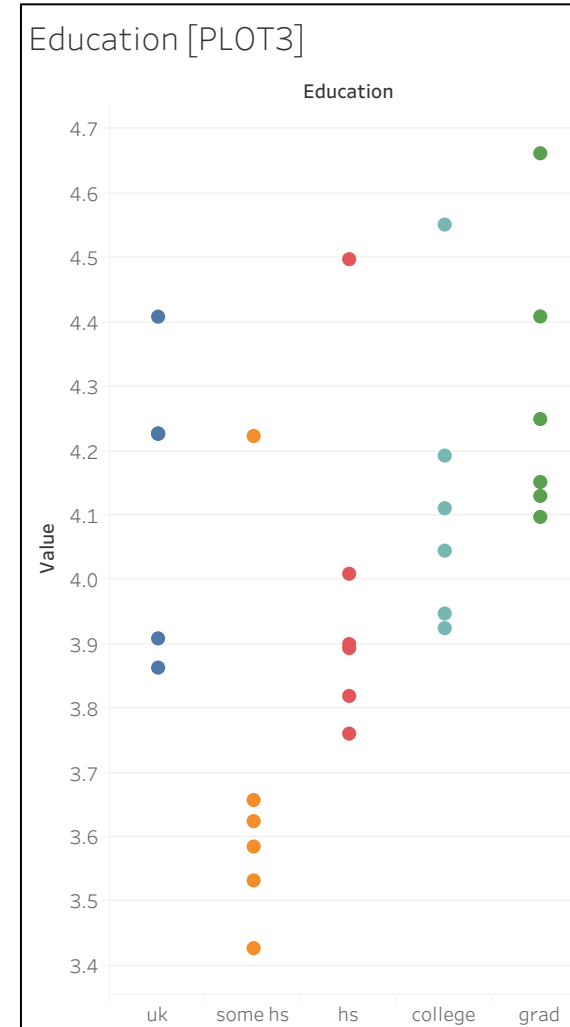
This plot is not poster-ready, but this is a very fast way to explore relationships in the data.



# Plot #3

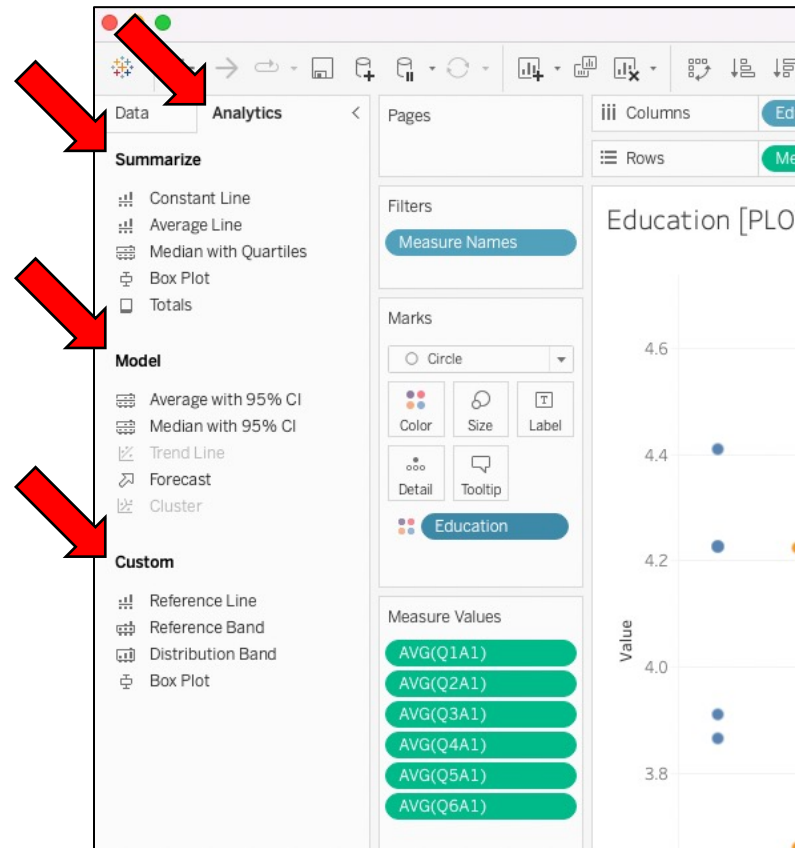
**Dot plots can be useful for showing distributions.**

Scatter plots are made the same way, but with 2 continuous variables. I'll show examples soon.



## Part 2: Next steps

# Analytics options (lines, boxplots, models)



**Analytics can take one of 3 scopes, depending on the variable types:**

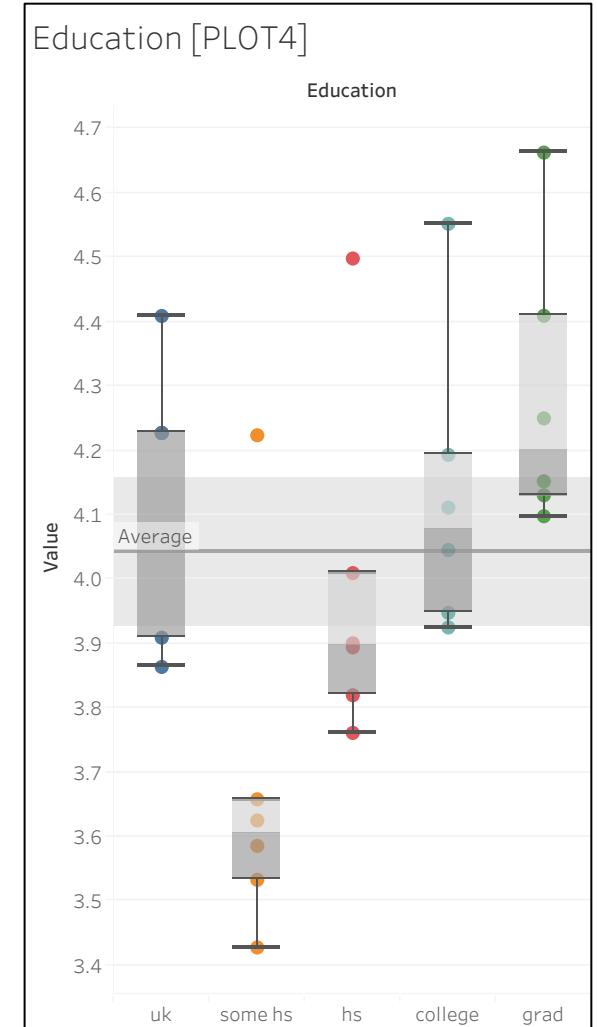
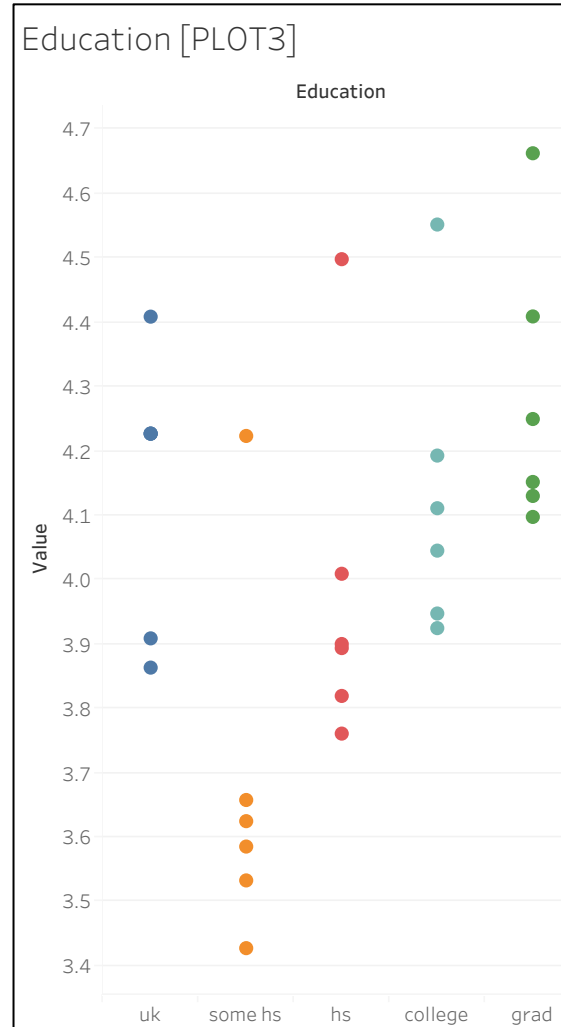
- a) cell-wise (i.e. averaging within participant)
- b) pane-wise (i.e. averaging within condition)
- c) table-wise (i.e. averaging across participants and conditions)

# Plot #4

**Adding boxplot overlays  
and an average line over  
the dot plot for clarity**

Tails & thresholds are  
straightforward to adjust by  
right-clicking a box and  
selecting Edit.

Hiding versus filtering “uk”  
–what is the difference?



# About Dataset #2

The data comes from a mobile app that screens for Autism Spectrum Disorder which were collected from an open online form at [openspsychometrics.com](http://openspsychometrics.com)



## About the Autism Tests App

This app is solely for research purposes and not for commercial use. The app contains Autism Spectrum Disorder (ASD) tests to enable parents, care givers, and academic researchers access them. These tests are not diagnostic tools rather they are behavioural tests that just pinpoint to autistic traits. The app can be freely downloaded and distributed but cannot be used for financial gain. This app was developed by Dr Fadi Fayeze at NMIT.



Thabtah, F. (2017). ASDTests. A mobile app for ASD screening.  
[www.asdtests.com](http://www.asdtests.com) [accessed December 20th, 2017]

# Quick look at Dataset #2

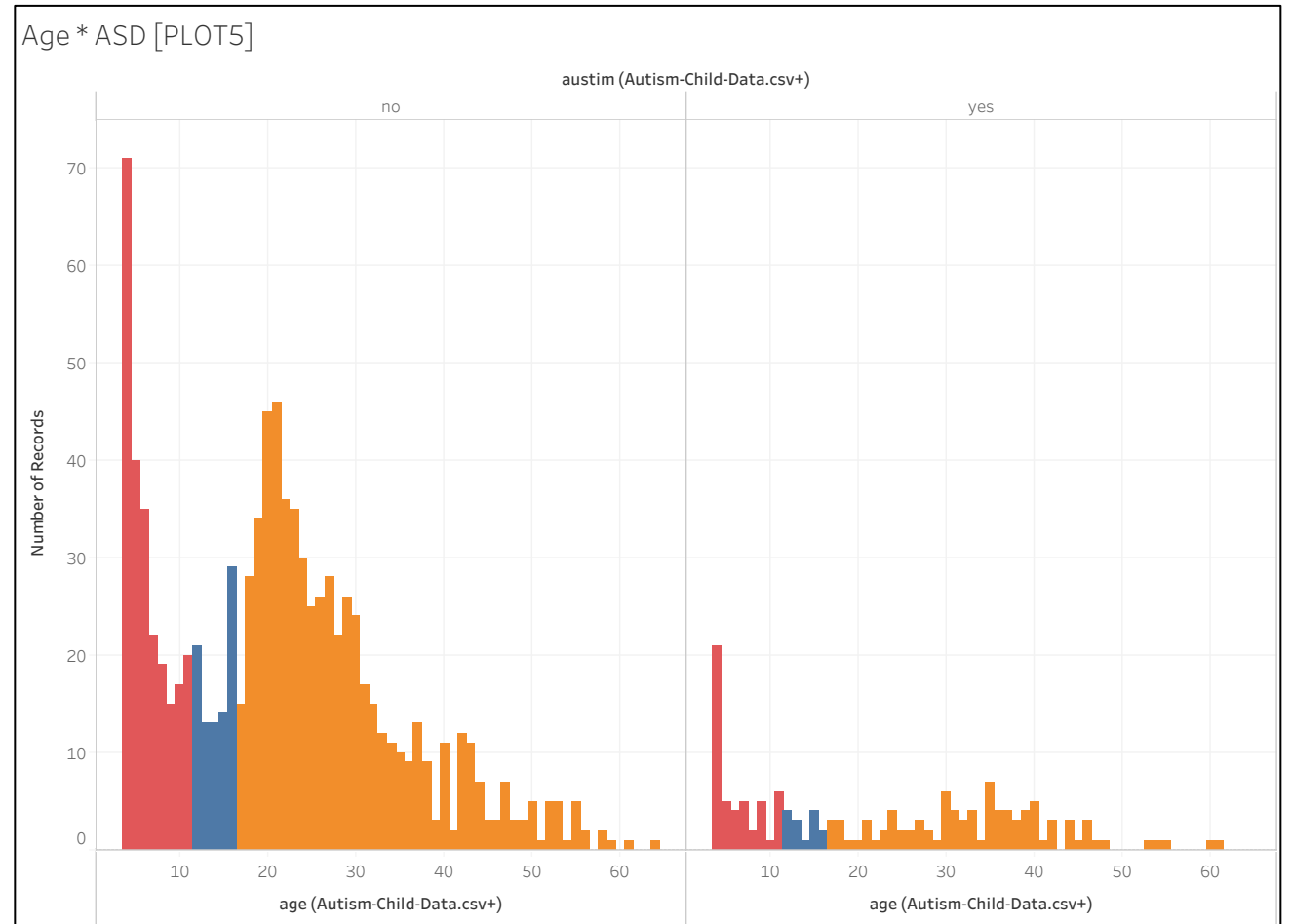
A1 _Score	A2 _Score	A3 _Score	A4 _Score	A5 _Score	age	gender	ethnicity	jundice	austim	contry_of_res	used_app_ before	result	age_desc	relation	Class/ASD
1	1	1	1	0	26	f	White-European	no	no	United States	no	6	18 and more	Self	NO
1	1	0	1	0	24	m	Latino	no	yes	Brazil	no	5	18 and more	Self	NO
1	1	0	1	1	27	m	Latino	yes	yes	Spain	no	8	18 and more	Parent	YES
1	1	0	1	0	35	f	White-European	no	yes	United States	no	6	18 and more	Self	NO

- This data comes packaged in 3 separate data files, 1 for each age group
- Connecting all 3 files in the Data Source pane lets us look at all of them

# Plot #5

**Histogram of ages, with two panels for a key variable.**

A key to success in Tableau is restraint. It's possible to make uninterpretable plots.

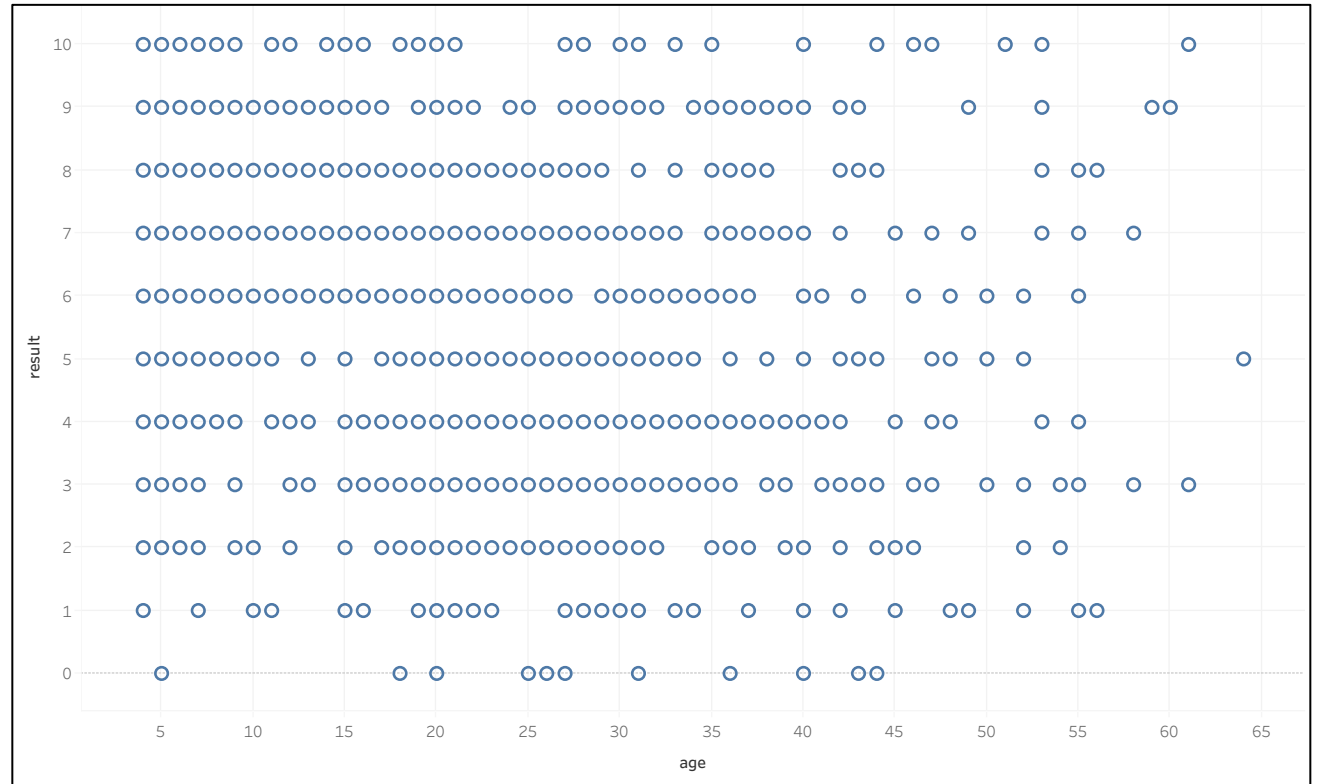




# Plot #6a

**A dot plot in need of the Analytics pane tools.**

Scatter plots are made the same way, but with 2 continuous variables. I'll show examples soon.

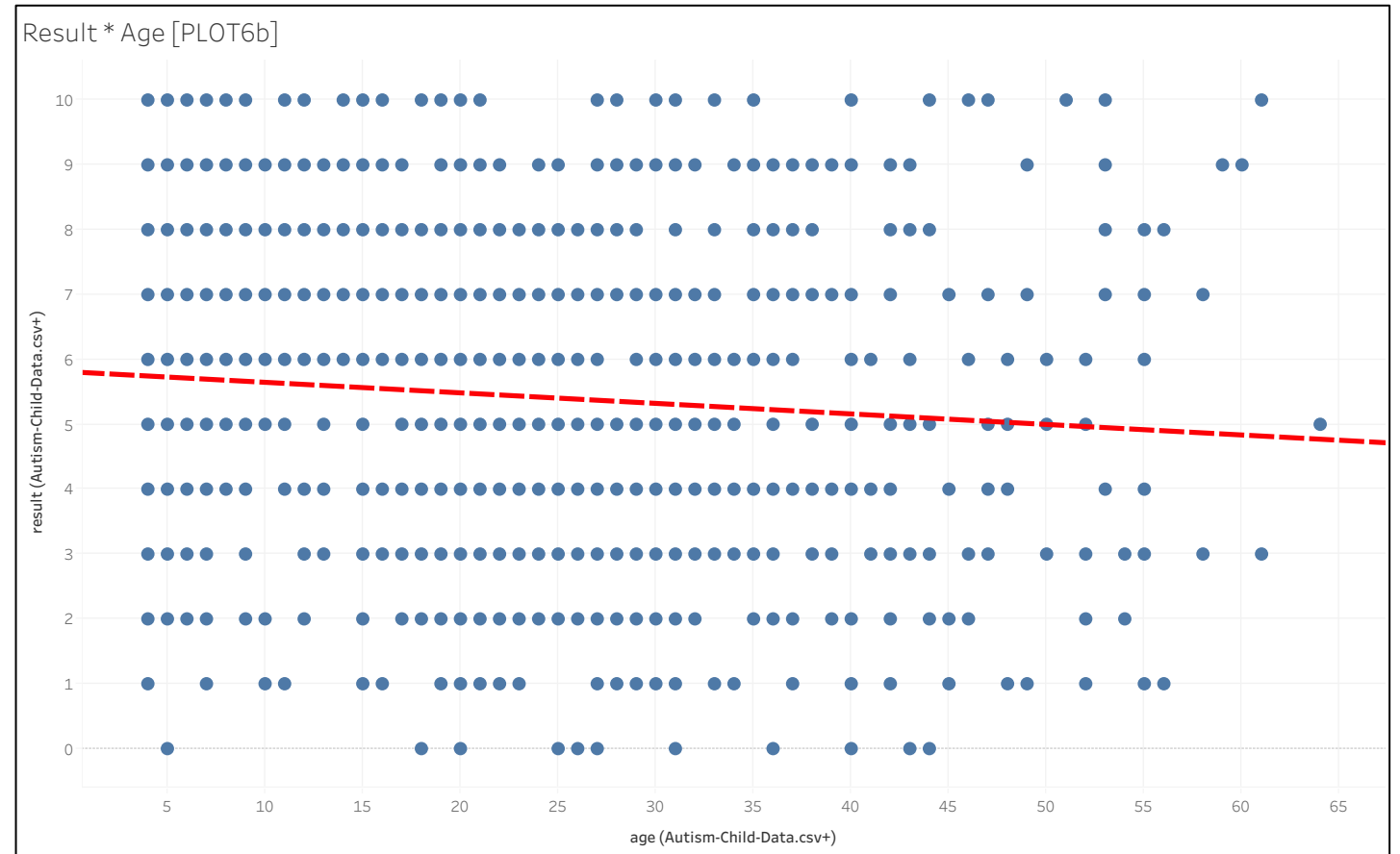
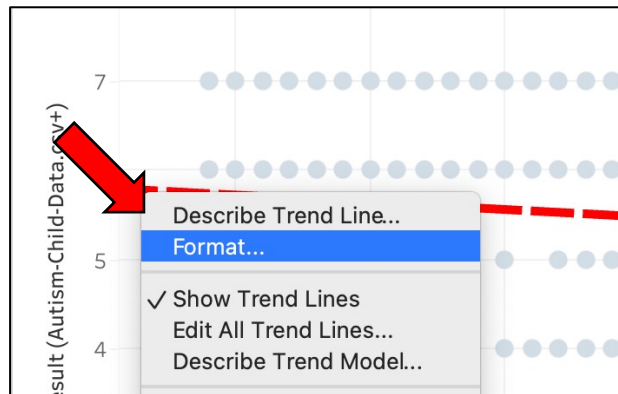


# Plot #6b

## Linear regression

Dragging “Trend Line” to “Linear” gives us a linear regression line. Mouse over to read p-values.

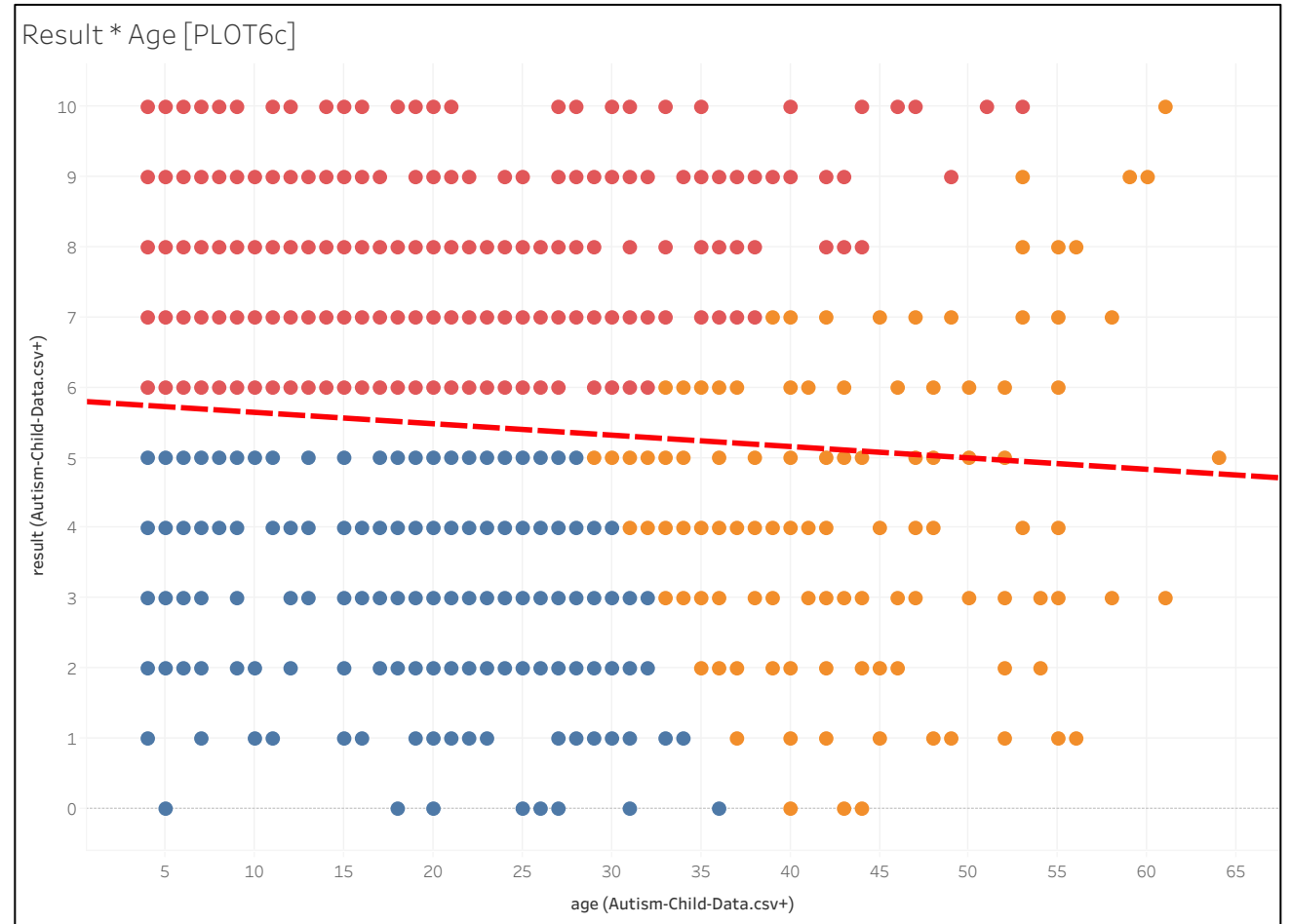
Tip: Right-click the line to Format it:



# Plot #6c

**Linear regression +  
K-means clustering**

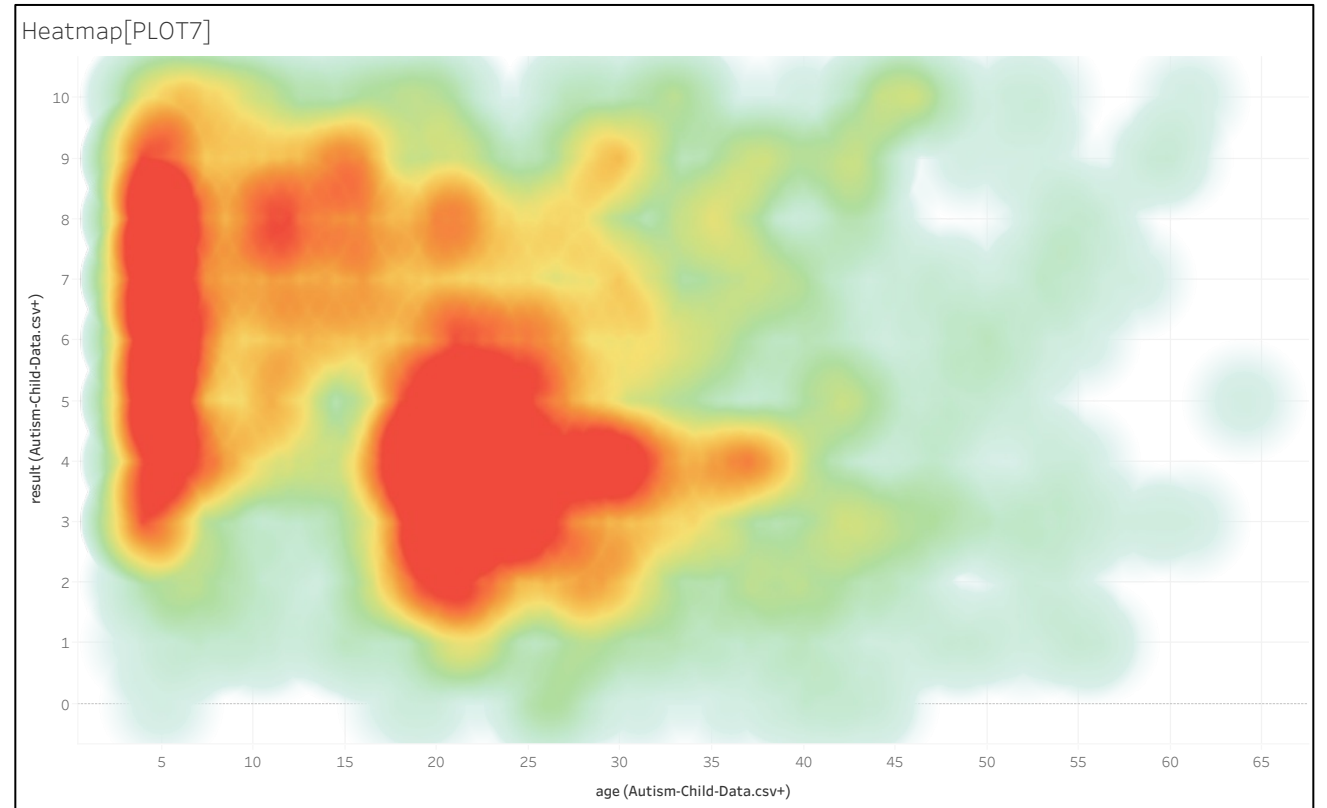
I don't think you can get  
clustering information  
out of Tableau easily.



# Plot #7

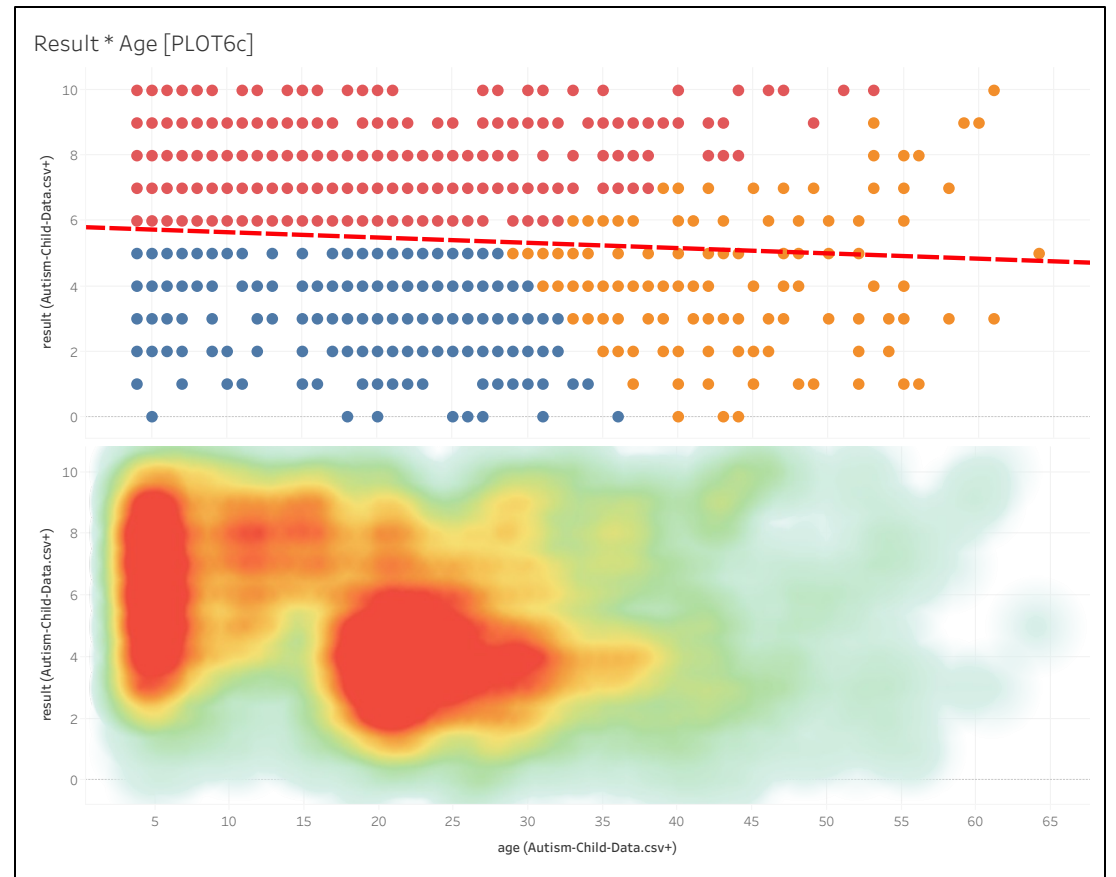
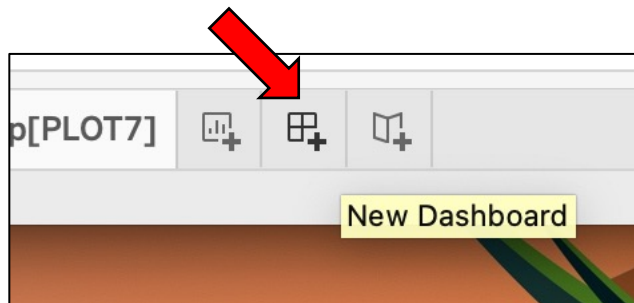
## Density plot (heatmap)

There seems to be 2 different patterns, earlier ASD (age 5-10) and later (20-35), which has been reported in other work in this area.



# How to create a Dashboard

- A Dashboard is an arrangement of Tableau Worksheets (figures)
- They are useful for sharing a set of figures with shared legends, subplots, and relating figures.



Simple Dashboard with 2 Worksheets

# Part 3: Case study

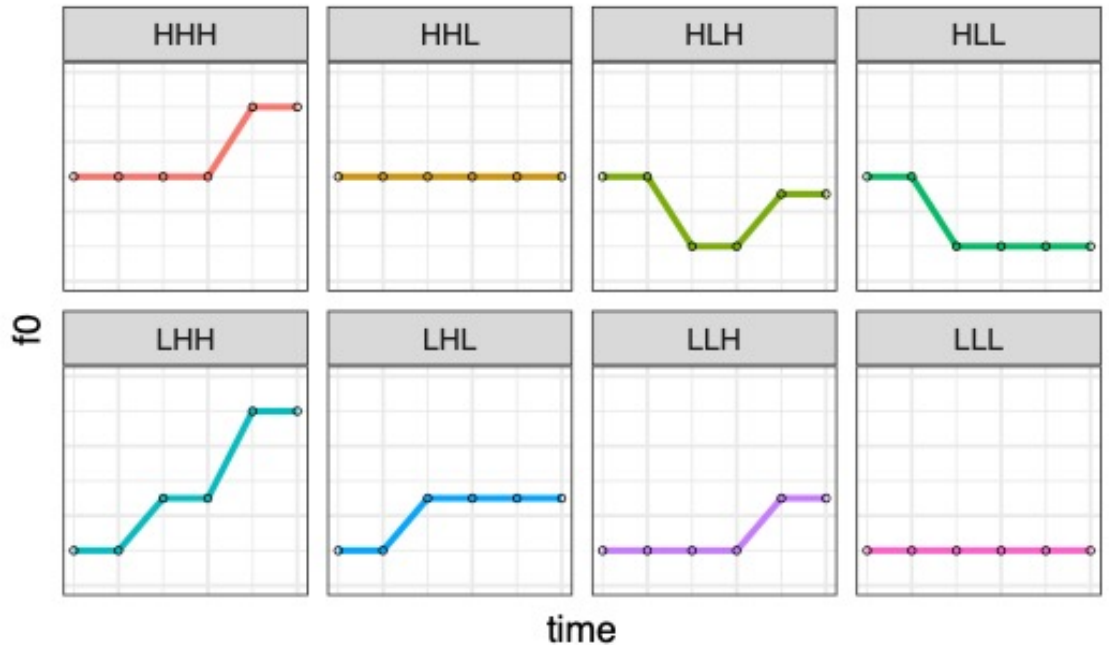
Looking at time series data from a recent study of mine...

# About Dataset #3

This data comes from a recent experiment that involved speakers imitating pitch patterns (variable = F0) in different emotional contexts.

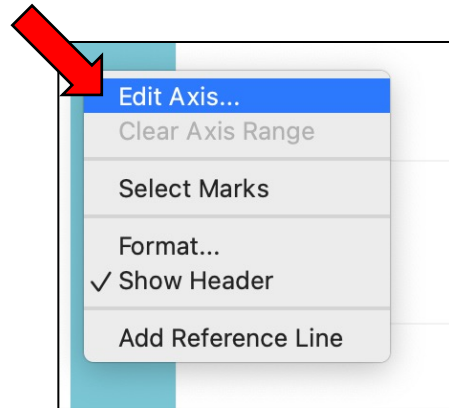
I took 10 equidistant time samples of F0 (pitch) per pattern (variable = Sample).

Line goes up, pitch goes up...



Turner, D. (2022). Intonational cues in emotional contexts. In prep. 😊

# Edit Axis

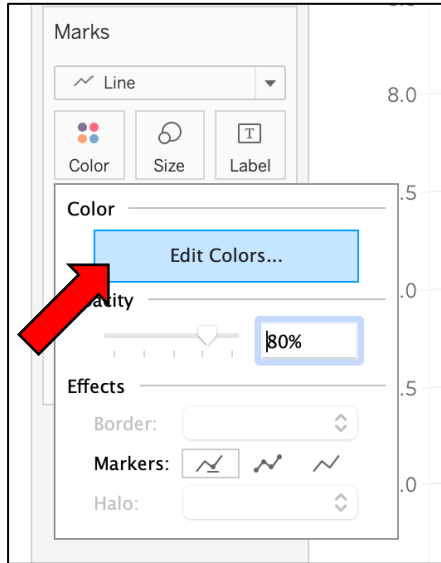


- Sometimes the values for your Measures will never cross zero, so it might not be informative to show it.
- From the Axis menu, you can choose to not 'Include zero', retitle axis, transform them, and choose whether each Panel in a figure should use the same or different scales.

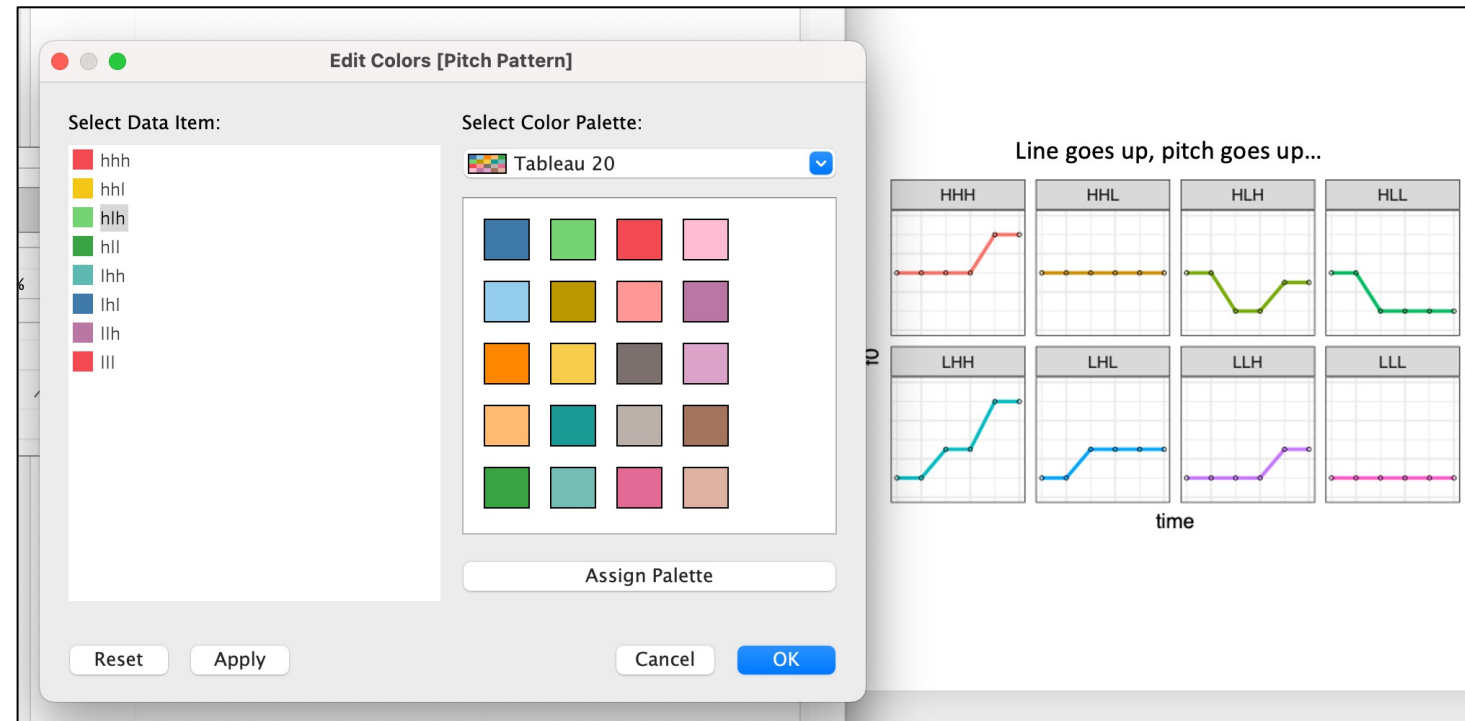
A screenshot of the 'Edit Axis [Avg. F0]' dialog box. The 'General' tab is selected. The 'Range' section has 'Automatic' selected with a radio button, and the 'Include zero' checkbox is checked. Below this, there are two input fields: the first contains '0' and the second contains '6.922493740'. The 'Scale' section has 'Positive' selected with a radio button. The 'Axis Titles' section has 'Title' set to 'Avg. F0' and 'Subtitle' set to 'Subtitle', with the 'Automatic' checkbox checked. A 'Reset' button is at the bottom left.



# Custom colors



- The schematized pitch patterns have 8 colors that were assigned when by ggplot2 in R.
- Tableau makes it easy to coordinate color legends.



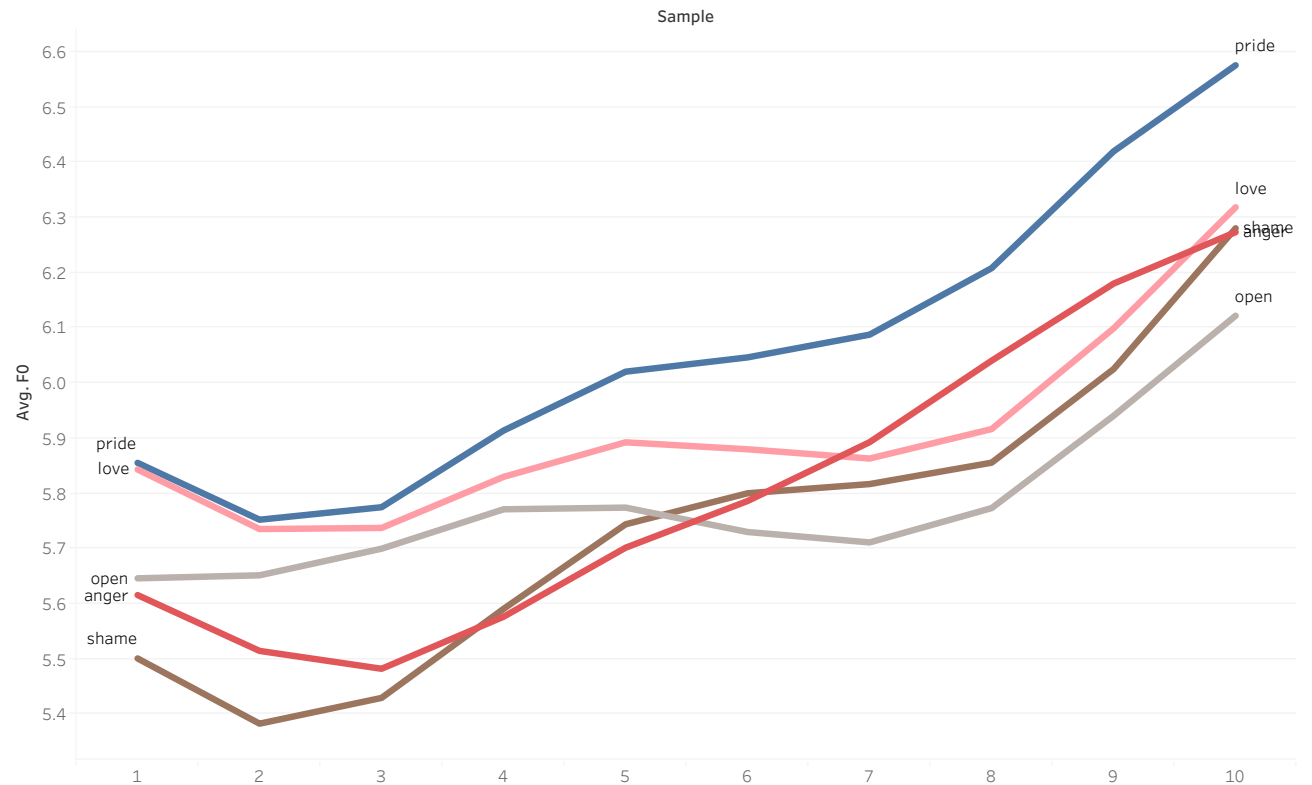
Recoding colors for these figures took less time than writing this caption.

# Plot #8

**Line plot showing 1 line per emotion condition with the average F0 for each time Sample.**

Tableau is automatically aggregating across participants, pitch pattern, and sample.

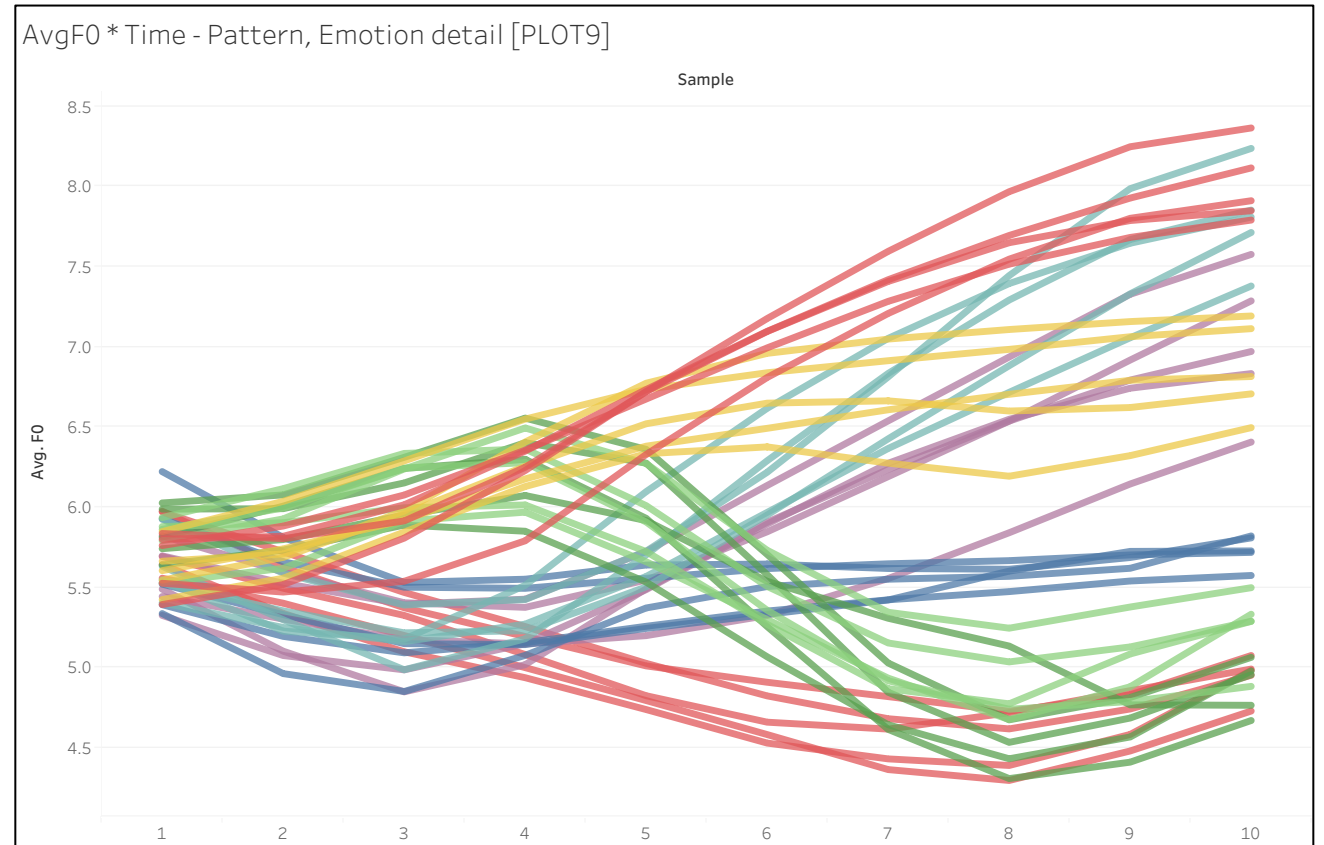
AvgF0 \* Time - Emotion [PLOT8]



# Plot #9

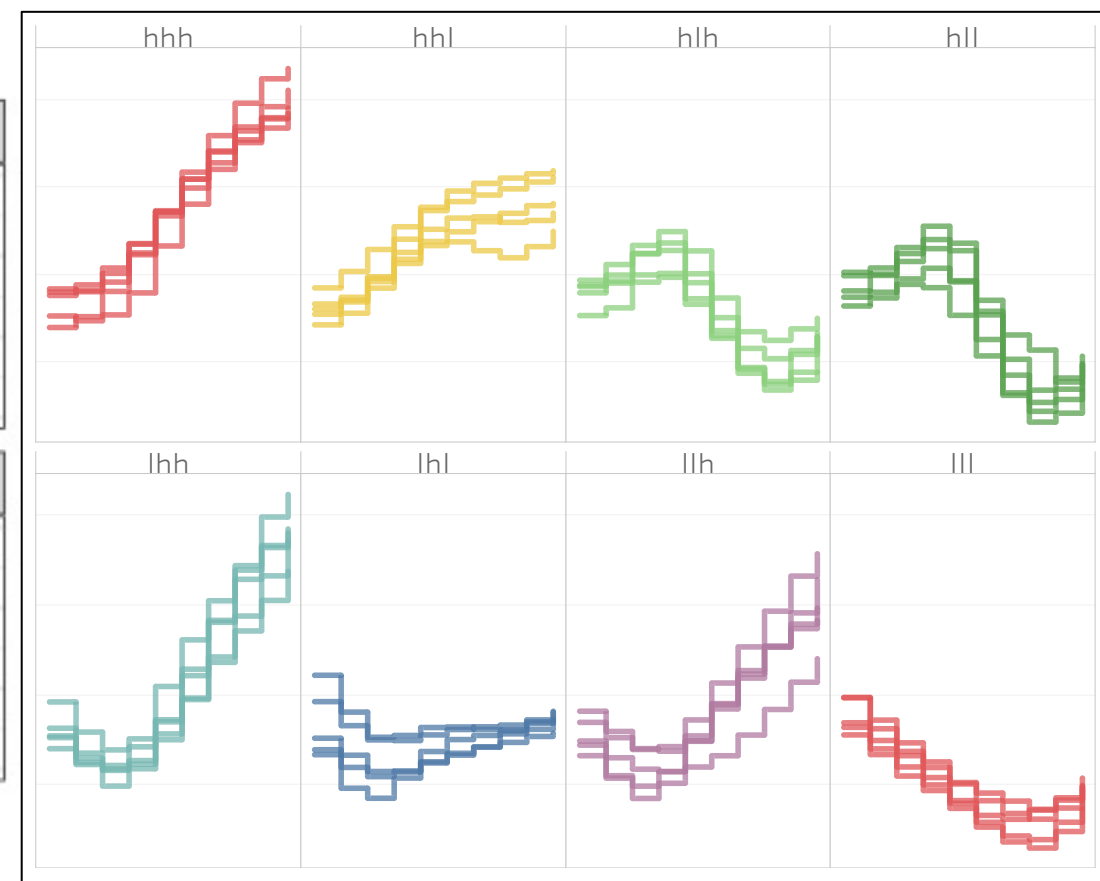
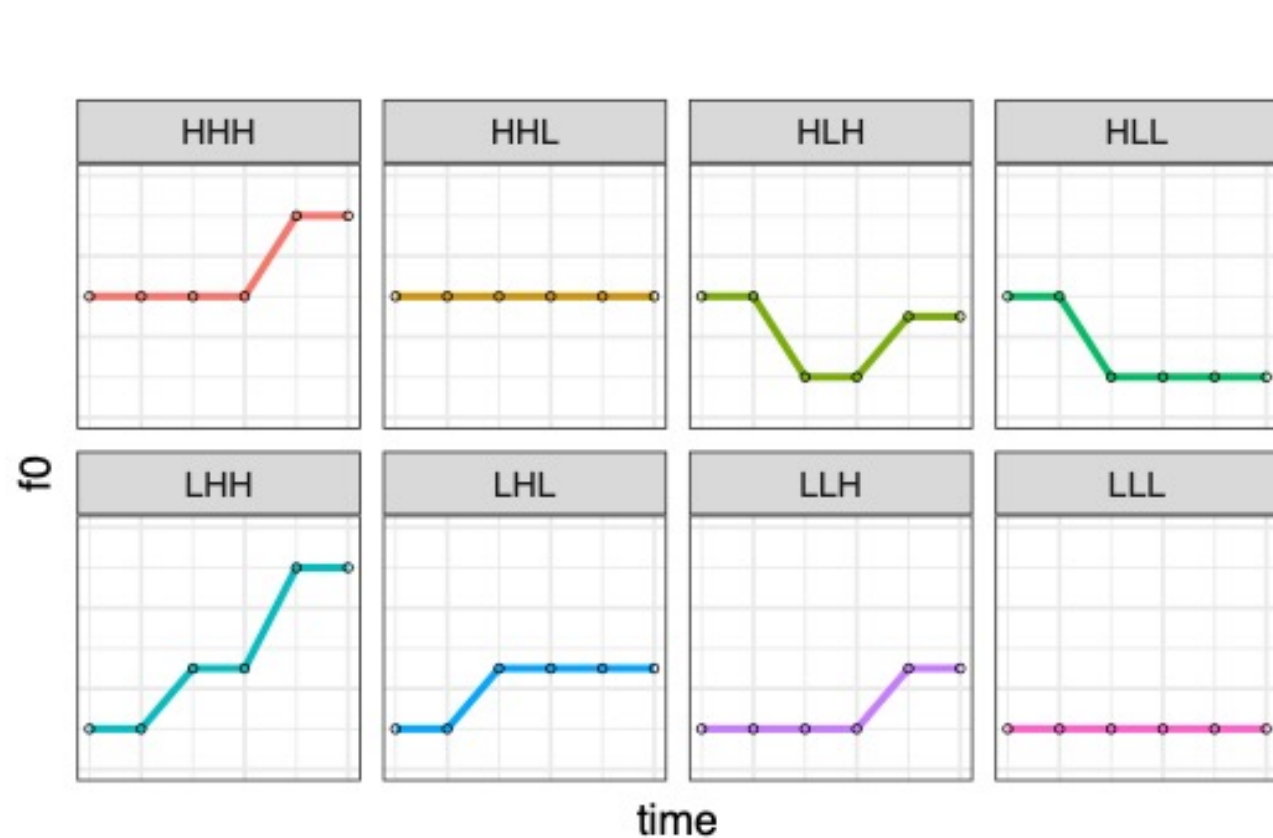
**Line plots of the 8 pitch patterns, with a separate line for each emotion too, so there are 5 lines for each pattern.**

This figure shows within-pitch pattern variation, which was my goal for this data.



# Plot #10

Comparing the schematized patterns to plots of the pitch measurements.



# Final Thoughts

- Tableau is a powerful tool for data visualization with some handy tools for analytics and working with many different types of data, BUT:
- R and Python (and friends) are with us forever, while software has a short lifespan. Plus, people can reproduce your findings if you publish a script that details every step of your analysis.
- Best to use Tableau for what it's good at:

**Making high quality figures of your raw data to help you explore, understand, and explain your data.**

Thank you!

[dturner@u.northwestern.edu](mailto:dturner@u.northwestern.edu)