Theme 4: Design Portfolio

Hayden Coffey and Ashwin Poduval

Stacked bars and positional encodings

The first set of designs described here were inspired by a mix of stacked bar charts and area charts

One axis in stacked bars is used for scalability/comparison

We decided to encode distribution information along that axis, retaining the other axis’s role in providing information about part-whole relationships

Our first set of designs merely encoded standard deviation using the thickness of the graph, as shown in the slide after the next one

Normalized vs Absolute standard deviation

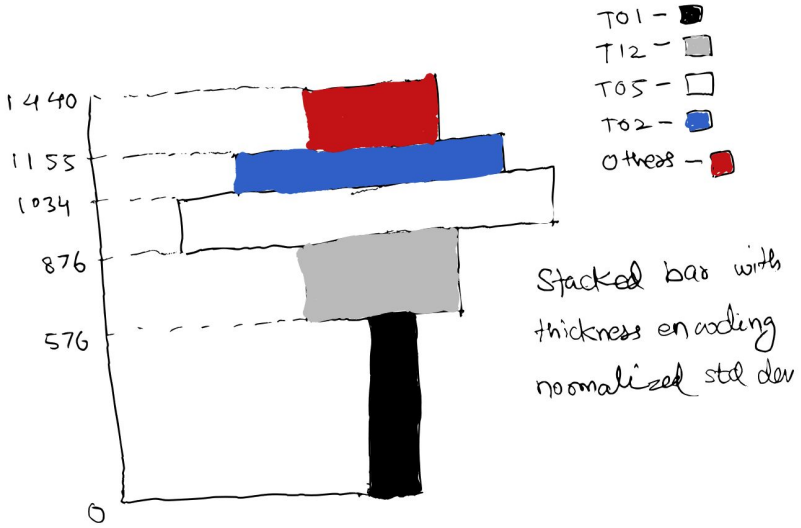
We realized that categories with higher means would be more likely have higher values of standard deviation

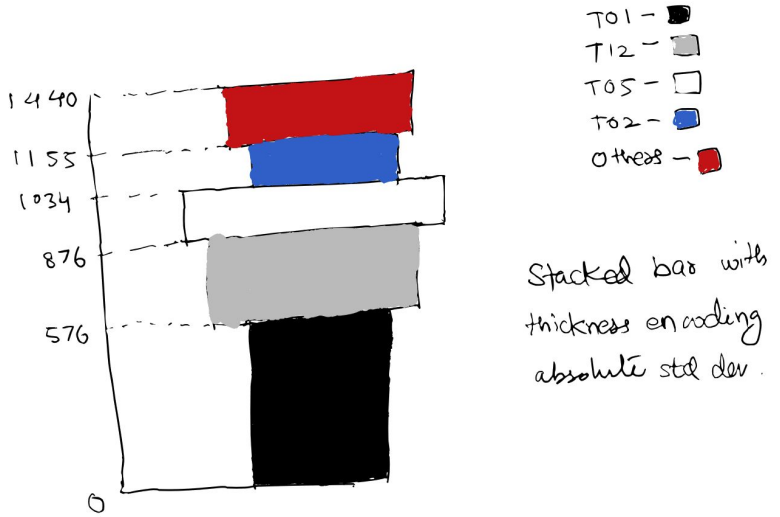
Thus, we chose to normalize the value of standard deviation by the mean to obtain a ratio helping us understand spread with respect to mean

For some of the hand-drawn figures, we grouped some of the tasks together to make it easier to create data accurate sketches; the stats are listed below

Category	Mean	Std. Dev./Normalized Std. Dev.
T01 (Personal Care)	about 576 min (40% of day)	143.3/0.25
T12 (Leisure, Socializing)	about 300 min (20-21% of day)	212.6/0.7
T05 (Work)	about 158 min (10-11% of day)	240.2/1.5
T02 (Household Activities)	about 121 min (8.4% of day)	141.1/1.17
All others	about 285 min (~20% of day)	190.9/0.67

Hand drawn sketches: stats used in table on previous slide





Making further use of the x-axis

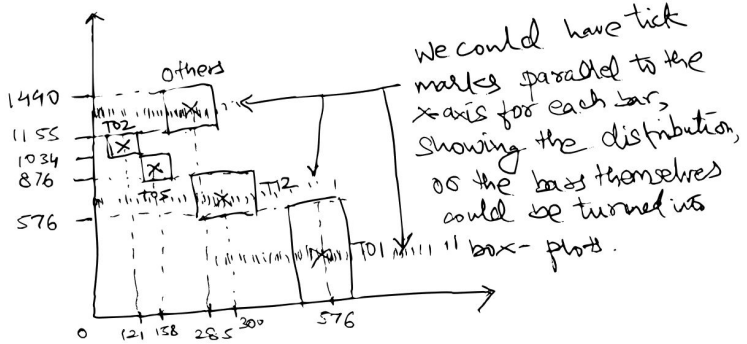
While information about standard deviation is helpful, it may not provide the full story to a viewer

Our next design used bars of uniform thickness, we placed category bars at the corresponding mean locations

To prevent overlap, we had the bars span differents heights

This was like a stacked bar chart, with displaced bars

We encoded distribution information using tick marks



Refinement

Placing bars at different locations hurt readability

Viewers eyes would have to jump around the figure

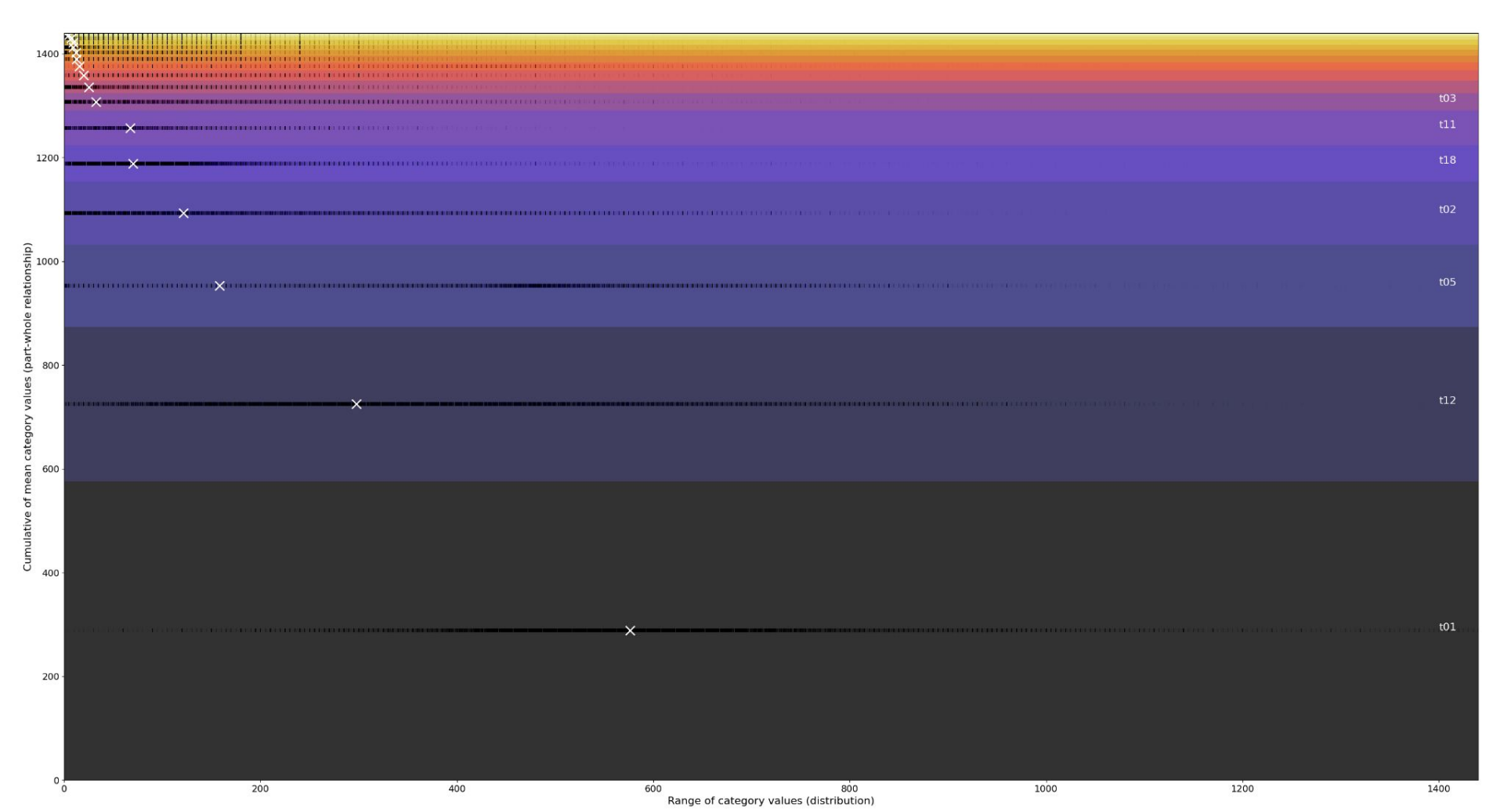
We decided to scrap the bars

We could stack areas parallel to the x-axis which respected the part-whole relation

Represented the distribution of the aggregated variable in this band

Iterated and developed an initial design, as shown in the next slide

Decided to implement it programmatically because drawing accurate hand-drawn figures was becoming an increasingly expensive proposition


t03
t11
t18
t02
t05
t12
t01
Cumulative of mean category values (part-whole relationship)
Range of category values (distribution)
0
200
400
600
800
1000
1200
1400

Problems with initial design

Some categories are bound to contribute to only a small fraction of the whole

We sorted the areas from bottom to top for readability, the largest contributors to the part whole are placed at the bottom, and the smallest at the top

We marked category mean values, shown by the white crosses

However, a bunch of smaller values ended up clustered in the top left

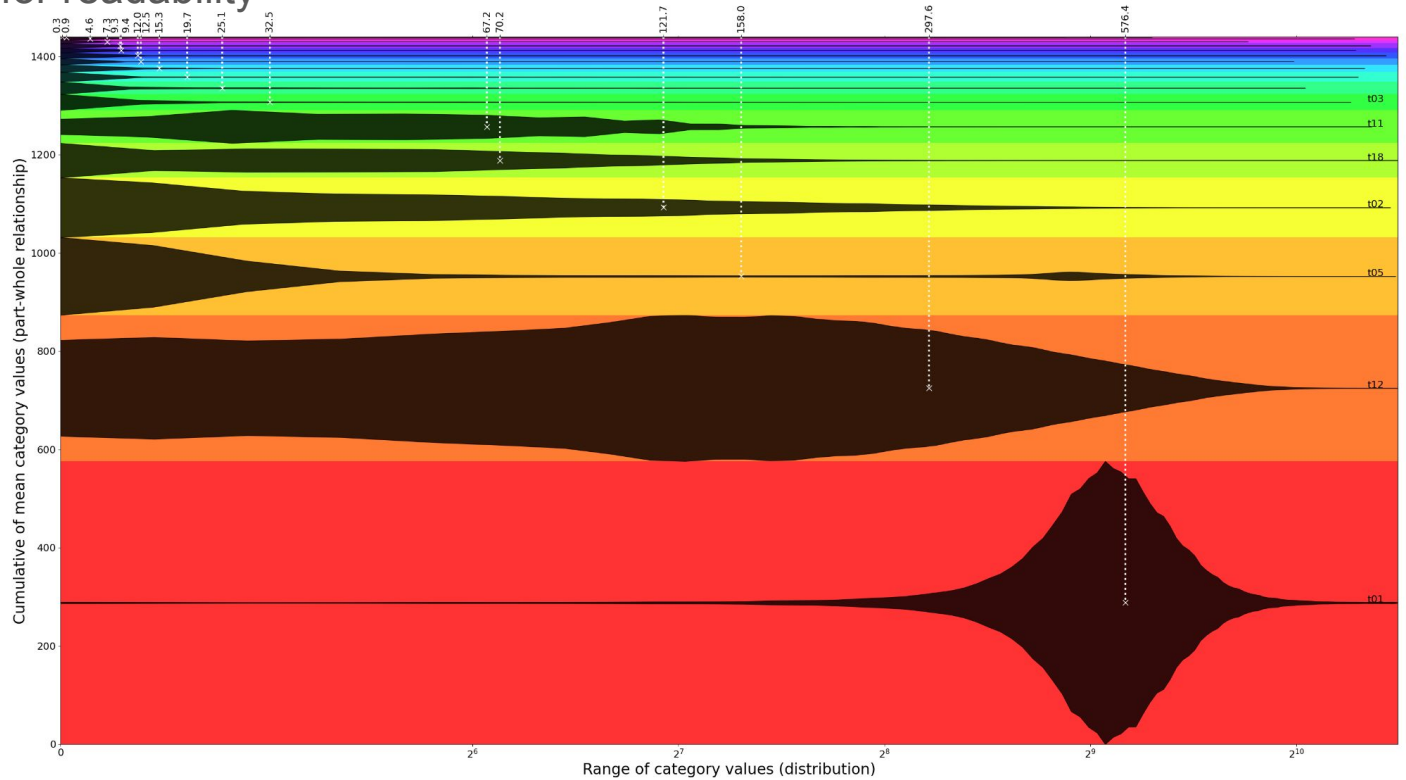
We represented distributions with bands of black tick marks parallel to x-axis

However, it was hard to tell if they were concentrated at a number, especially if they were at 0, since they would overlap with the y-axis

Shifted to a semilog scale & used violin plots to see if it made things better

Changed color encoding - weren't sure if the colors previously implied some order

Added dashed lines from means to top of fig with readings and marked category text on the far right, for readability



Combining some of the best aspects of both

Scrapped colors - used spanning arrows and category codes to indicate regions

Used linear scale to avoid distorting violin plots

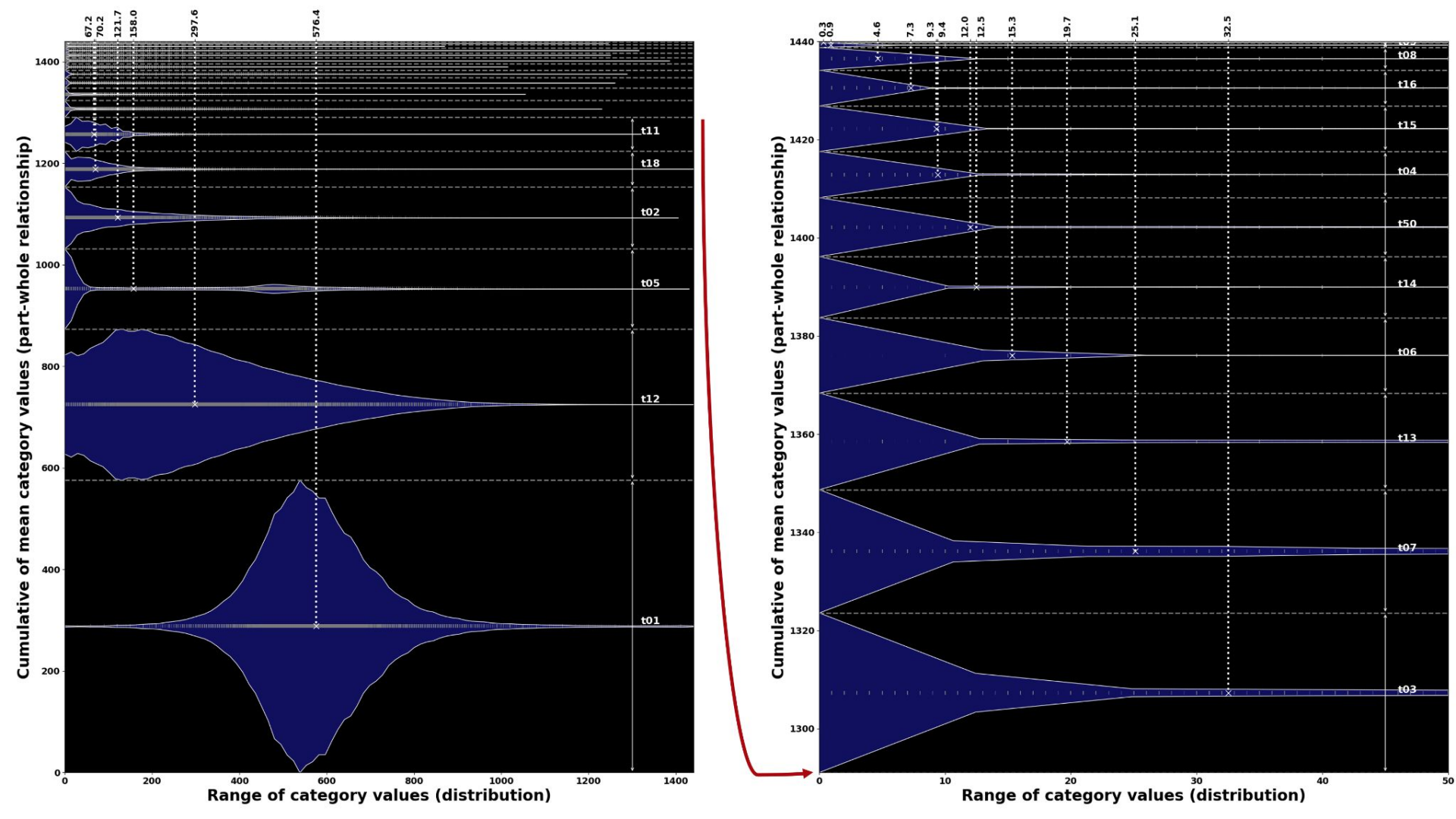
Added violin plots & tick marks to give users max information about distribution

Used subplot to zoom into region of unreadable elements in part-whole

In this case, 12 categories contributed to about 150 minutes out of 1440

These were mostly unreadable

Interesting takeaways from the graph - we can clearly see that T05 has a bimodal distribution of sorts, and its mean is between both peaks


67.2
70.2
121.7
158.0
297.6
576.4
t11
t18
t02
t05
t12
t01
Cumulative of mean category values (part-whole relationship)
Range of category values (distribution)
0.3
0.9
4.6
7.3
9.3
9.4
12.0
12.5
15.3
19.7
25.1
32.5
t08
t16
t15
t04
t50
t14
t06
t13
t07
t03
Cumulative of mean category values (part-whole relationship)
Range of category values (distribution)

More effective part-whole representation

Even with the sub-plot, some of the smallest elements were indistinguishable

One might argue that such categories may not be interesting given their miniscule contribution

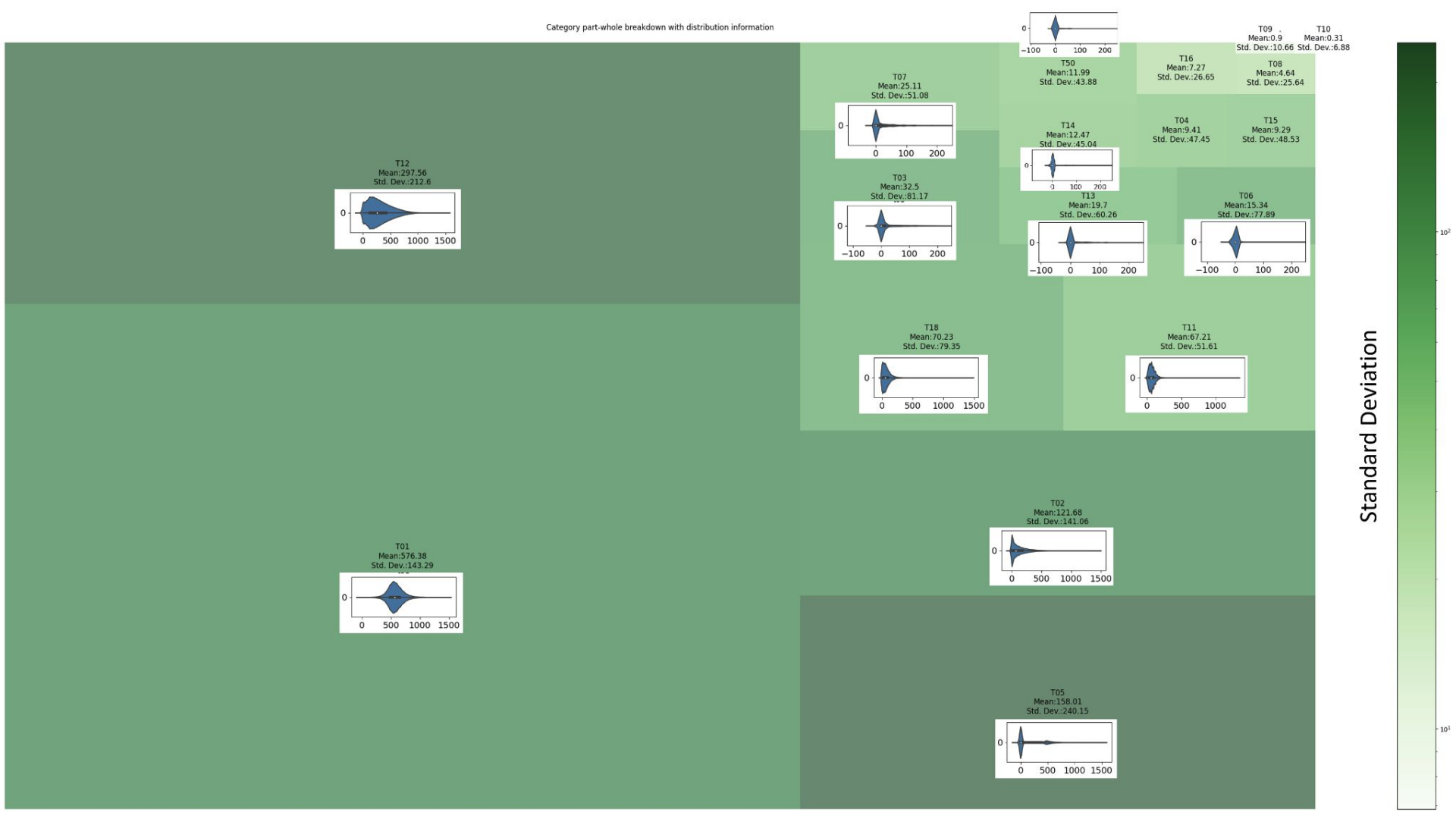
We tested area encodings to see if they could help draw attention to all members

We used treemaps, with standard deviation encoded with color

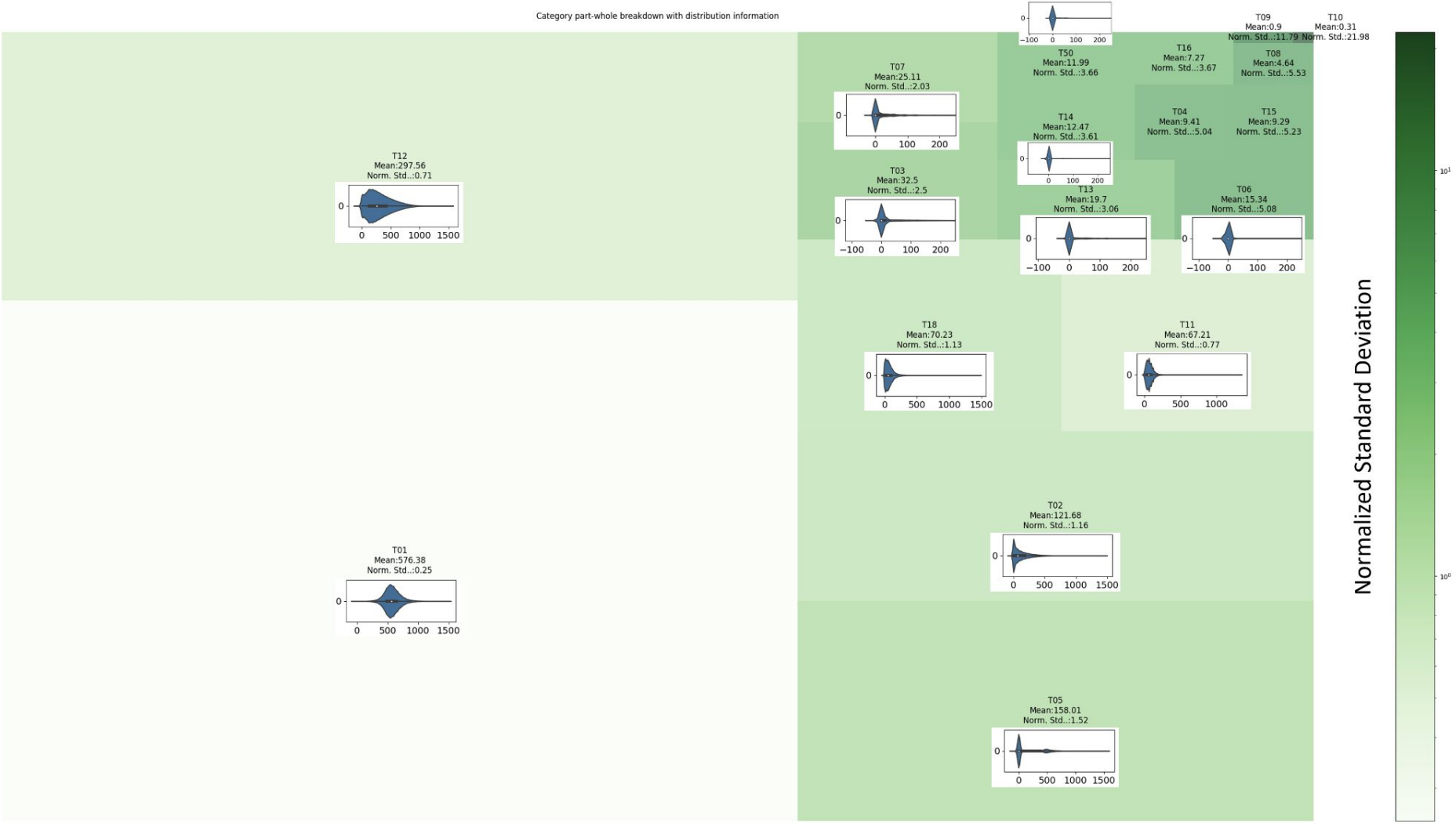
Provided exact mean and std. dev. values as text inside each block

Added violin plots to as many blocks as possible for distribution information

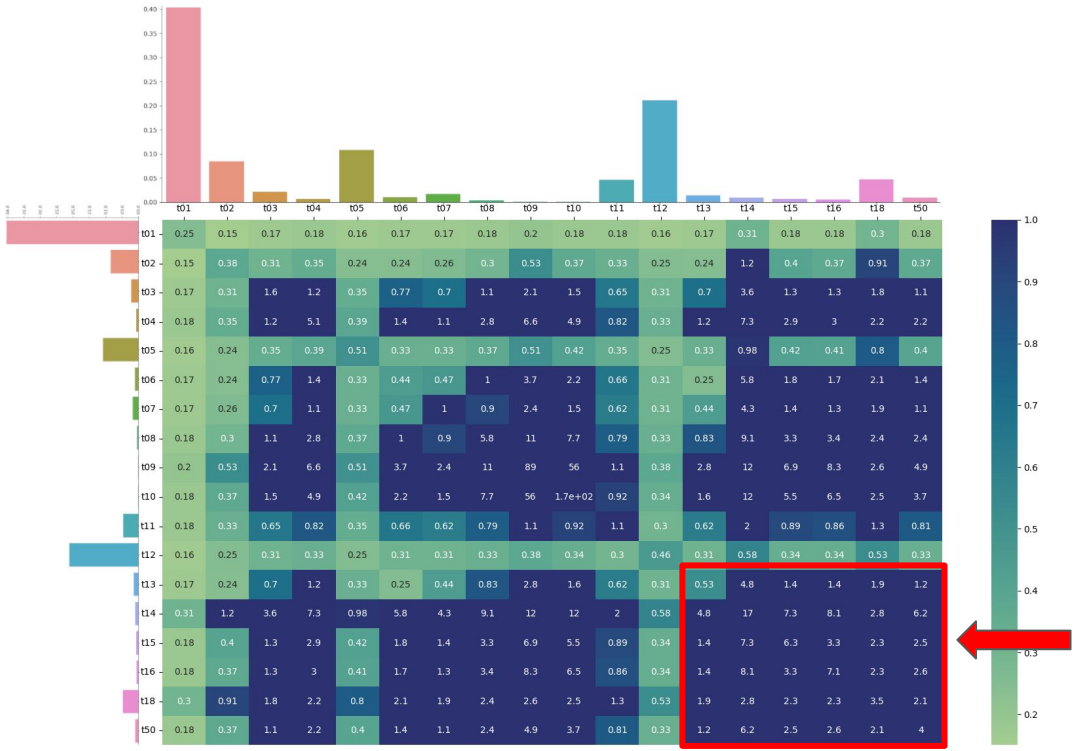
Made treemaps for both absolute and normalized standard deviation


Category part-whole breakdown with distribution information
T12
Mean:297.56
Std. Dev.:212.6
T07
Mean:25.11
Std. Dev.:51.08
T50
Mean:11.99
Std. Dev.:43.88
T09
Mean:0.9
Std. Dev.:10.66
T10
Mean:0.31
Std. Dev.:6.88
T16
Mean:7.27
Std. Dev.:26.65
T08
Mean:4.64
Std. Dev.:25.64
T14
Mean:12.47
Std. Dev.:45.04
T04
Mean:9.41
Std. Dev.:47.45
T15
Mean:9.29
Std. Dev.:48.53
T03
Mean:32.5
Std. Dev.:81.17
T13
Mean:19.7
Std. Dev.:60.26
T06
Mean:15.34
Std. Dev.:77.89
T18
Mean:70.23
Std. Dev.:79.35
T11
Mean:67.21
Std. Dev.:51.61
T02
Mean:121.68
Std. Dev.:141.06
T01
Mean:576.38
Std. Dev.:143.29
T05
Mean:158.01
Std. Dev.:240.15
Standard Deviation

Category part-whole breakdown with distribution information

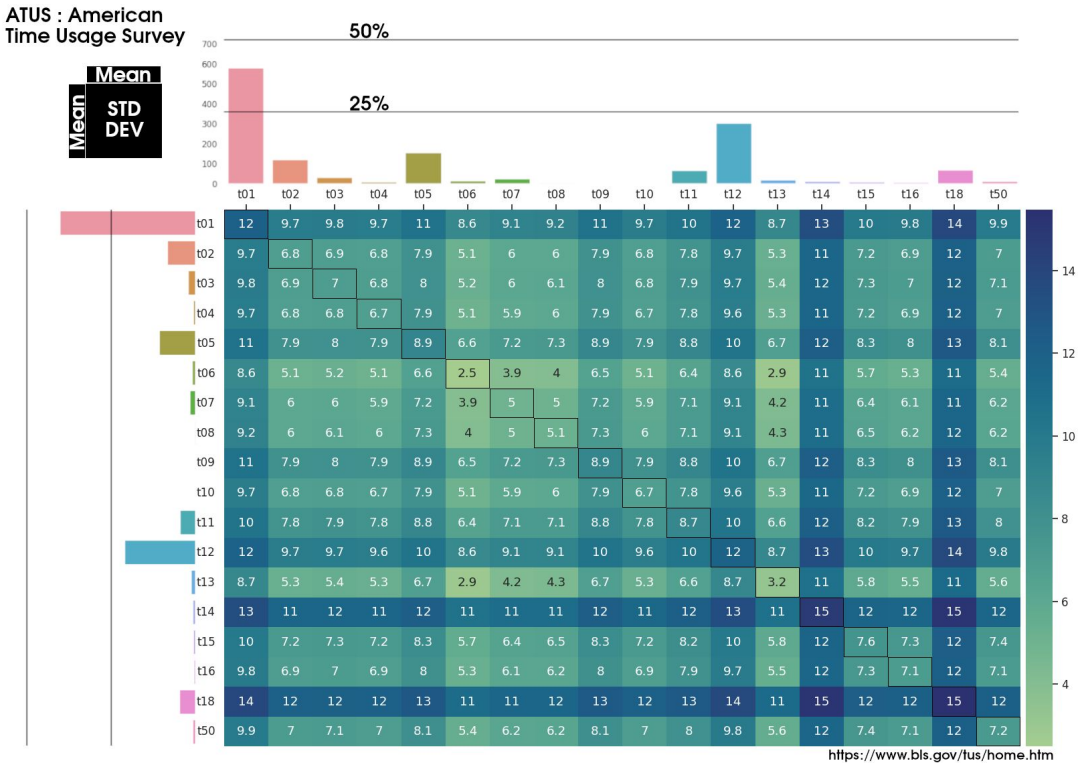


Heatmap Rough Draft: Standard Dev Encoded as %

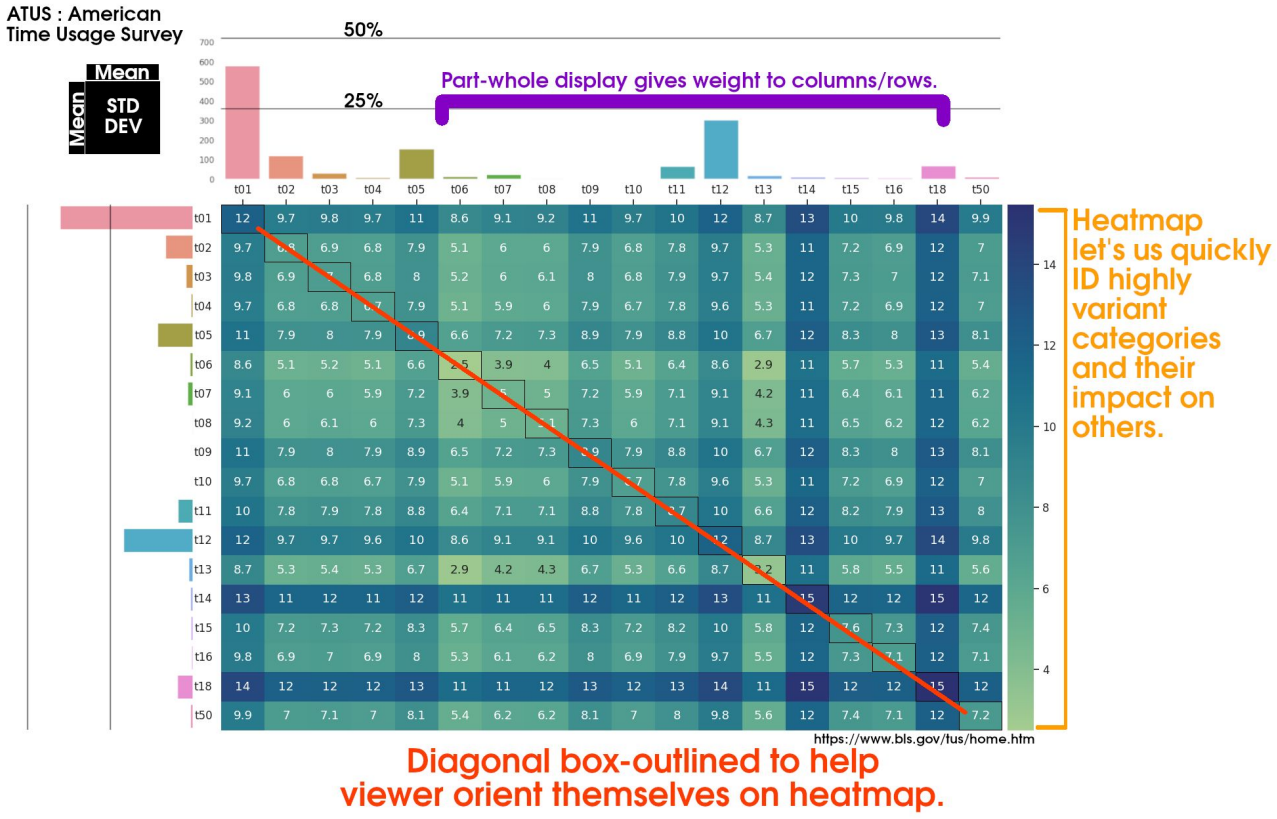


Too many maxed out cells.

Heatmap Final : Present Actual Standard Dev, Not %



Heatmap Final: Design Overview

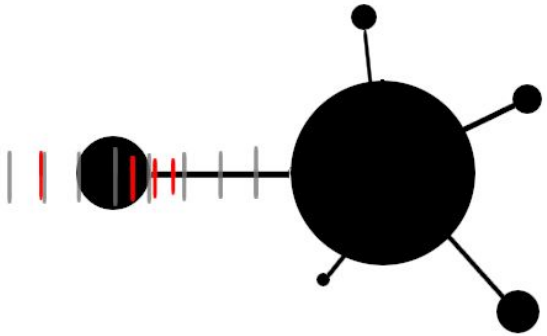


Node-Link: Rough Draft

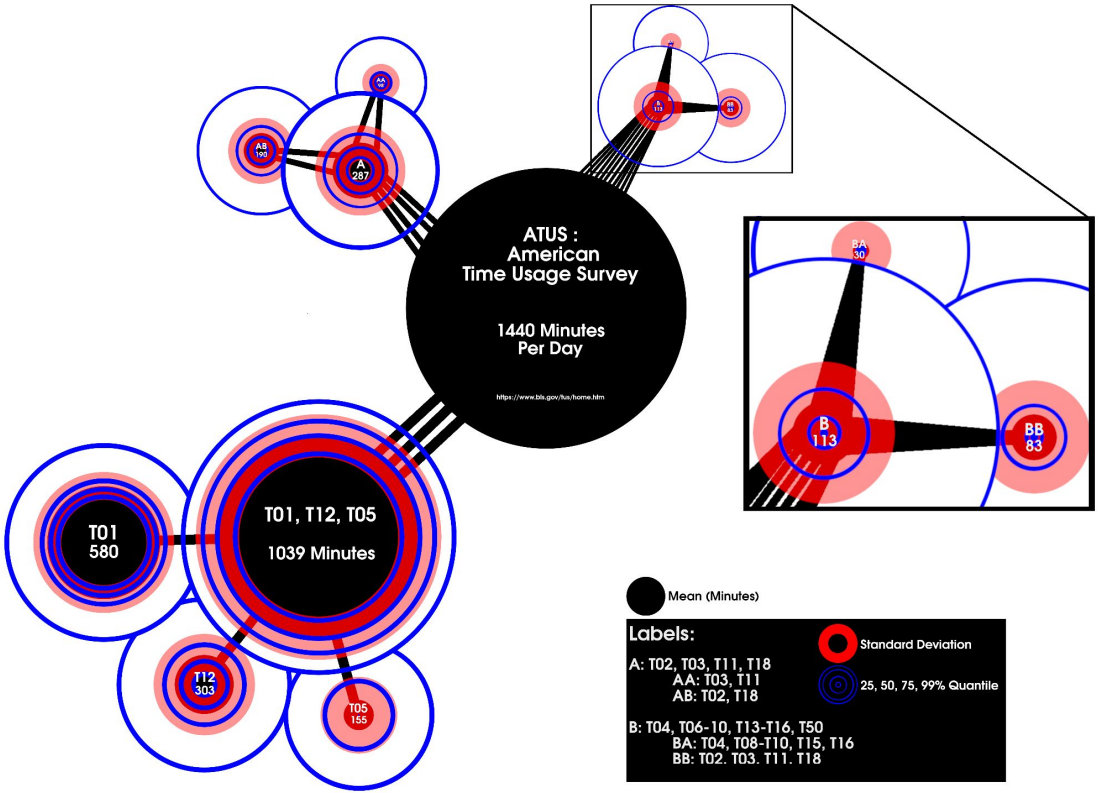
Size encode mean with node.

Grey ticks provide sense of scale.

Red ticks capture quantiles.



Node-Link: Final



Node-Link: Final Design Overview