Problems with American Community Survey

Dan Price

2022-01-09

Preliminaries – very broad

A few years ago, a small group of us decided to create a pedagogical tool for future health professionals. We wanted to give these students something idealized enough to be easy to work with and real enough to be engaging and relevant to their future work lives. Talk about the social determinants of health and neighborhood effects on health outcomes had recently become quite pervasive, and we thought that a better understanding of how the existing health professionals were modeling health outcomes within neighborhood contexts would be valuable and would arise naturally from their engagement in grounded and practical work with data. We had found it frustrating, however, to deal with the data that was published and publicly available, and so decided to simulate a county-size population for the students, where they could compare their local knowledge with what the data was showing. In that effort to create a simulated Harris County - Sam City, a more friendly version of Houston, which is mostly included in Harris County - we tried to work with the American Community Survey (ACS) from the US Census Bureau. The decennial census is broken into small groups and gives the answers provided by respondents to the household questionnaires in block groups. In the intervening years, the ACS provides estimates for the same categories, as well as producing a few new tables, although we soon found there were several important differences in how they thought of their competing priorities when producing the data tables overall. They had concerns about data privacy, statistical fidelity, and the broad utility of the data tables themselves, each of which provided both challenges for our attempt and - we thought - interesting pedagogical opportunities for our students. We thought it would be helpful (and, somewhat naively, easy) to quickly show where we were running into difficulties and why.

One of our tasks, as teachers, was to make clear what the motivating questions were around the collection and representation of data, and the ACS was also a very good fit for the organizing metaphors that drove the discussion of the social determinants of health. A big part of the idea that was capturing people's attention, after all, is that things one knows about a place can somehow provide insight into how health outcomes are structured by the opportunities, constraints, and daily habits of communal life. We knew that there were limits to this way of framing the overall questions - the concentration on health outcomes, to speak directly to the most important problem, tends to reduce the complicated pathways individuals travel and the variety of ongoing and changing processes we associate with robust good health to a constrained number of measurable states that an individual can be in relative to medical paradigms of treatment and billing. We knew it would be somewhat complicated, but having the students understand how everything fit together was promising, because the complications pointed to ways that our future health professionals could make effective interventions. Our hope was that our students could eventually use those methods to intervene in outcomes and make a difference in people's health, both at individual and community levels.

We knew of (and taught to our students) Nancy Krieger's critique of the metaphor of the web of causation in "Epidemiology and the Web of Causation: Has Anyone Seen the Spider" (https://www.sciencedirect.com/science/article/abs/pii/027795369490202X), and wanted our students to see the methods of epidemiology, no matter what theory of methodology they were developed within, to be constituted by a set of choices that they could understand and manipulate. Her work had powerfully criticized the idea, implicit in the notion of the web of causation, that there are individual threads of causation that one can modify in isolation and she had convinced us that the still dominant mode of looking for multiple causation using multifactorial analysis had deep political roots in the cold war's conception of individualism. She framed her work very specifically as a question of how epidemiology should be taught at the graduate level, and began with specific textbooks that had used the metaphor to justify and encourage "cutting strands rather than attempting to identify and alter the source of the web" (p. 890). By contracst, she wanted to ask, as her title implied, what the causes behind the building of the web as a whole were, and not just whether or not an individual in isolation could be freed from its adverse effects.

She proposed ecosocial theory as an alternative framing, or organizing image of the whole. Borrowing popular metaphors from Stephen Jay Gould in terms of ecology and evolution, she made clear that her task was not to simply identify the "spiders" - or root causes - but to find a better organizing image for the collective work of professional epidemiologists. She did not see herself as returning to a model that eliminated "irreducible individuality" in the name of an abstract analysis of systems that supposedly produce and determine what it means to be an individual, but as having suggested a model for the interaction between the individual and that context that would provide a grounded, more nuanced, and more complete understanding of how that individuality was channeled and shaped. It's worth quoting at length, since we saw ourselves as trying to support this turn toward a different organizing image, but will be suggesting some different directions for exploring alternatives.

"It is of little help to posit that health and disease are socially produced within evolving and socially-conditioned biologic parameters without offering insight into why and how this occurs; reducing the 'spiders' to a new form of 'black box' would only reinforce existing limitations. Nor would introducing the the 'spiders' necessarily reolve the 'web's' embodiment of a biomedical and individualistic worldview. The 'web' never was intended to and does not jar epidemiologists from the long-established practice of viewing population patterns of disease as simply the sum of individual cases; it is far from obvious that adding the 'spiders' would address the fundamental problem. ¶ As an alternative, the closest image that comes to mind stems from marrying the metaphor of the continually-constructed 'scaffolding' of society that different social groups daily seek to reinforce or alter to that of the ever-growing 'bush' of evolution, together defining the potential and constraints of human life. The intertwining ensemble must be understood to exist at every level, sub-cellular to societal, repeating indefinitely, like a fractal object. Different epidemiologic profiles at the population-level would accordingly be seen as reflecting the interlinked and diverse patterns of exposure and susceptibility that are brought into play by the dynamic intertwining of these changing forms. It is an image that does *not* permit the cleavage of the social from the biologic, and the biologic from the social. It is an image that does not obscure agency. And it is an image that embraces history rather than hides it from view." (p. 896).

Our idea was to put the students into the place of understanding some of the limitations of the way that the data was collected - the history of census questions on race and ethnicity, the absence of social class and income data in biomedical tables, the mundane problems of aggregation and representation - and to have them understand that much of their agency as health professionals rested in ensuring that their work was responding to the metaphors that they were consciously imposing. Then, for example, the work on the census data should be approached not as a mere counting exercise, providing the pool from which individual cases are tallied and individual interventions are measured, but as the scaffolding within which individual trajectories are constrained or encouraged.

Could talk about the computational problem, and more specifically, the sudoku vs. rubic's cube question. Both of these are ways of dealing computationally with the scaffolding and bush problem. Could also talk about the difficulty of actually cashing out the "fractal" metaphor.

We began this process several years before the 2020 census had been completed, although as of this writing the first data has been published and we will talk about how to integrate it later. Our first task was to see whether the ACS estimates could help us. Block group data is the smallest unit made available to the public and is made available for some ACS tables, but we ran into the problem at the tract level, which is comprised of from 1 to 4 block groups, and the block group totals aggregate correctly into the tracts. (https://www.census.gov/programs-surveys/geography/about/glossary.html) Using our libraries (Census_Data.R and workflow.R) we were able to save the appropriate data locally. I reproduce the code, below, for completeness' sake, but there's no need to follow it closely. Using the three tracts that we had selected because we were familiar with the areas and they represented demographic variability, we wanted to look at the distribution of females by age, and then by race. When aggregating by tract. For the 2010 decennial census, these are women by age (with some overlap in age categories).

```
dec_sex_by_age_race_data_from_census_10 <-
      censusData_byGroupName(censusdir, vintage="2010", state, censuskey,
                             groupname = "P12",county_num = "201",
                             block="block_group",api_type="dec/sf1",path_suff="est.csv")
```

```
## [1] "found folder ~/Downloads/UH_OneDrive/OneDrive - University Of Houston/Social Network Hy
## [1] "Reading file from ~/Downloads/UH_OneDrive/OneDrive - University Of Houston/Social Netwo
```

```
SAR_2010 <- as.data.table(dec_sex_by_age_race_data_from_census_10)
#all the designations have o in them except totals for all races
F_SAR_2010 <- SAR_2010[str_detect(label,"Female")&!str_detect(concept,"O")]
F3_SAR_2010 <- F_SAR_2010[order(label),
                  list(`label`,`48_201_312200_1`,`48_201_312200_2`,`48_201_312200_3`,
                       `48_201_310300_1`,`48_201_310300_2`,`48_201_310300_3`,
                       `48_201_310300_4`,`48_201_310300_5`,`48_201_310300_6`,
                       `48_201_411900_1`,`48_201_411900_2`,`48_201_411900_3`)]
#add summary columns
F3_SAR_2010$`48201312200` <- as.integer(F3_SAR_2010$`48_201_312200_1`) +
  as.integer(F3_SAR_2010$`48_201_312200_2`) +
  as.integer(F3_SAR_2010$`48_201_312200_3`)
F3_SAR_2010$`48201310300` <- as.integer(F3_SAR_2010$`48_201_310300_1`) +
  as.integer(F3_SAR_2010$`48_201_310300_2`) +
  as.integer(F3_SAR_2010$`48_201_310300_3`) +
```

```r
    as.integer(F3_SAR_2010$`48_201_310300_4`) +
    as.integer(F3_SAR_2010$`48_201_310300_5`) +
    as.integer(F3_SAR_2010$`48_201_310300_6`)
  F3_SAR_2010$`48201411900` <- as.integer(F3_SAR_2010$`48_201_411900_1`) +
    as.integer(F3_SAR_2010$`48_201_411900_2`) +
    as.integer(F3_SAR_2010$`48_201_411900_3`)

  F3s_SAR_2010 <- F3_SAR_2010[order(label),
                  list(`label`,`48201312200`,`48201310300`,`48201411900`)]

F3s_SAR_2010
```

```
##                               label 48201312200 48201310300 48201411900
##  1:                   Total!!Female        1019        2274        1773
##  2:    Total!!Female!!10 to 14 years          61         124          58
##  3:    Total!!Female!!15 to 17 years          45          75          35
##  4:   Total!!Female!!18 and 19 years          40          54          13
##  5:          Total!!Female!!20 years          12          35          12
##  6:          Total!!Female!!21 years          12          25          18
##  7:    Total!!Female!!22 to 24 years          45         122         120
##  8:    Total!!Female!!25 to 29 years          66         233         182
##  9:    Total!!Female!!30 to 34 years          65         188         151
## 10:    Total!!Female!!35 to 39 years          55         156         124
## 11:    Total!!Female!!40 to 44 years          71         144         105
## 12:    Total!!Female!!45 to 49 years          69         138         108
## 13:      Total!!Female!!5 to 9 years          63         147          69
## 14:    Total!!Female!!50 to 54 years          90         165         153
## 15:    Total!!Female!!55 to 59 years          57         162         167
## 16:   Total!!Female!!60 and 61 years          24          36          58
## 17:    Total!!Female!!62 to 64 years          29          59          82
## 18:   Total!!Female!!65 and 66 years          19          50          49
## 19:    Total!!Female!!67 to 69 years          25          43          59
## 20:    Total!!Female!!70 to 74 years          23          58          49
## 21:    Total!!Female!!75 to 79 years          27          36          33
## 22:    Total!!Female!!80 to 84 years          13          28          27
## 23: Total!!Female!!85 years and over          26          32          34
## 24:     Total!!Female!!Under 5 years          82         164          67
##                               label 48201312200 48201310300 48201411900
```

If we further restrict the display to Black females

```r
  BF_SAR_2010 <- SAR_2010[str_detect(label,"Female")&str_detect(concept,"BLACK")]
  BF3_SAR_2010 <- BF_SAR_2010[order(label),
                  list(`label`,`concept`,`48_201_312200_1`,`48_201_312200_2`,
                       `48_201_312200_3`,
                    `48_201_310300_1`,`48_201_310300_2`,`48_201_310300_3`,
```

```r
                        `48_201_310300_4`,`48_201_310300_5`,`48_201_310300_6`,
                        `48_201_411900_1`,`48_201_411900_2`,`48_201_411900_3`)]
    #add summary columns
    BF3_SAR_2010$`48201312200` <- as.integer(BF3_SAR_2010$`48_201_312200_1`) +
      as.integer(BF3_SAR_2010$`48_201_312200_2`) +
      as.integer(BF3_SAR_2010$`48_201_312200_3`)
    BF3_SAR_2010$`48201310300` <- as.integer(BF3_SAR_2010$`48_201_310300_1`) +
      as.integer(BF3_SAR_2010$`48_201_310300_2`) +
      as.integer(BF3_SAR_2010$`48_201_310300_3`) +
      as.integer(BF3_SAR_2010$`48_201_310300_4`) +
      as.integer(BF3_SAR_2010$`48_201_310300_5`) +
      as.integer(BF3_SAR_2010$`48_201_310300_6`)
    BF3_SAR_2010$`48201411900` <- as.integer(BF3_SAR_2010$`48_201_411900_1`) +
      as.integer(BF3_SAR_2010$`48_201_411900_2`) +
      as.integer(BF3_SAR_2010$`48_201_411900_3`)

    BF3s_SAR_2010 <- BF3_SAR_2010[order(label),
                      list(`label`,`concept`,`48201312200`,`48201310300`,`48201411900`)]
BF3s_SAR_2010
```

```
##                                label
##  1:                    Total!!Female
##  2:    Total!!Female!!10 to 14 years
##  3:    Total!!Female!!15 to 17 years
##  4:   Total!!Female!!18 and 19 years
##  5:          Total!!Female!!20 years
##  6:          Total!!Female!!21 years
##  7:    Total!!Female!!22 to 24 years
##  8:    Total!!Female!!25 to 29 years
##  9:    Total!!Female!!30 to 34 years
## 10:    Total!!Female!!35 to 39 years
## 11:    Total!!Female!!40 to 44 years
## 12:    Total!!Female!!45 to 49 years
## 13:      Total!!Female!!5 to 9 years
## 14:    Total!!Female!!50 to 54 years
## 15:    Total!!Female!!55 to 59 years
## 16:   Total!!Female!!60 and 61 years
## 17:    Total!!Female!!62 to 64 years
## 18:   Total!!Female!!65 and 66 years
## 19:    Total!!Female!!67 to 69 years
## 20:    Total!!Female!!70 to 74 years
## 21:    Total!!Female!!75 to 79 years
## 22:    Total!!Female!!80 to 84 years
## 23: Total!!Female!!85 years and over
## 24:     Total!!Female!!Under 5 years
##                                label
##                                          concept 48201312200 48201310300
```

```
##  1: SEX BY AGE (BLACK OR AFRICAN AMERICAN ALONE)         814          93
##  2: SEX BY AGE (BLACK OR AFRICAN AMERICAN ALONE)          47           4
##  3: SEX BY AGE (BLACK OR AFRICAN AMERICAN ALONE)          30           5
##  4: SEX BY AGE (BLACK OR AFRICAN AMERICAN ALONE)          28           1
##  5: SEX BY AGE (BLACK OR AFRICAN AMERICAN ALONE)          11           1
##  6: SEX BY AGE (BLACK OR AFRICAN AMERICAN ALONE)          10           3
##  7: SEX BY AGE (BLACK OR AFRICAN AMERICAN ALONE)          31           2
##  8: SEX BY AGE (BLACK OR AFRICAN AMERICAN ALONE)          47           8
##  9: SEX BY AGE (BLACK OR AFRICAN AMERICAN ALONE)          42           8
## 10: SEX BY AGE (BLACK OR AFRICAN AMERICAN ALONE)          35          12
## 11: SEX BY AGE (BLACK OR AFRICAN AMERICAN ALONE)          61           7
## 12: SEX BY AGE (BLACK OR AFRICAN AMERICAN ALONE)          60          13
## 13: SEX BY AGE (BLACK OR AFRICAN AMERICAN ALONE)          40           4
## 14: SEX BY AGE (BLACK OR AFRICAN AMERICAN ALONE)          83           8
## 15: SEX BY AGE (BLACK OR AFRICAN AMERICAN ALONE)          51           5
## 16: SEX BY AGE (BLACK OR AFRICAN AMERICAN ALONE)          23           0
## 17: SEX BY AGE (BLACK OR AFRICAN AMERICAN ALONE)          25           1
## 18: SEX BY AGE (BLACK OR AFRICAN AMERICAN ALONE)          18           1
## 19: SEX BY AGE (BLACK OR AFRICAN AMERICAN ALONE)          25           0
## 20: SEX BY AGE (BLACK OR AFRICAN AMERICAN ALONE)          21           2
## 21: SEX BY AGE (BLACK OR AFRICAN AMERICAN ALONE)          27           1
## 22: SEX BY AGE (BLACK OR AFRICAN AMERICAN ALONE)          13           1
## 23: SEX BY AGE (BLACK OR AFRICAN AMERICAN ALONE)          26           3
## 24: SEX BY AGE (BLACK OR AFRICAN AMERICAN ALONE)          60           3
##                                           concept 48201312200 48201310300
##     48201411900
##  1:          32
##  2:           0
##  3:           0
##  4:           0
##  5:           0
##  6:           0
##  7:           1
##  8:           7
##  9:           6
## 10:           3
## 11:           4
## 12:           2
## 13:           1
## 14:           2
## 15:           1
## 16:           2
## 17:           1
## 18:           0
## 19:           2
## 20:           0
## 21:           0
## 22:           0
```

```
## 23:          0
## 24:          0
##    48201411900
```

But if we look at the 2017 estimates at the tract level

```
    sex_by_age_race_data_from_census_17 <-
          censusData_byGroupName(censusdir, vintage, state, censuskey,
                                groupname = "B01001",county_num = "201",
                                block="tract",api_type="acs/acs5",path_suff="est.csv")
```

```
## [1] "found folder ~/Downloads/UH_OneDrive/OneDrive - University Of Houston/Social Network Hy
## [1] "Reading file from ~/Downloads/UH_OneDrive/OneDrive - University Of Houston/Social Netwo
```

```
    SAR_2017 <- as.data.table(sex_by_age_race_data_from_census_17)
    #all the designations have o in them except totals for all races
    F_SAR_2017 <- SAR_2017[str_detect(label,"Female")&!str_detect(concept,"O")]
    F3_SAR_2017 <- F_SAR_2017[order(label),
            list(`label`,`48201312200`,`48201310300`,`48201411900`)]
F3_SAR_2017
```

```
##                                            label 48201312200 48201310300
##  1:                   Estimate!!Total:!!Female:         908        2212
##  2:    Estimate!!Total:!!Female:!!10 to 14 years           1         133
##  3:    Estimate!!Total:!!Female:!!15 to 17 years          11          36
##  4:   Estimate!!Total:!!Female:!!18 and 19 years          24          28
##  5:          Estimate!!Total:!!Female:!!20 years          41           0
##  6:          Estimate!!Total:!!Female:!!21 years          11          20
##  7:    Estimate!!Total:!!Female:!!22 to 24 years          43         117
##  8:    Estimate!!Total:!!Female:!!25 to 29 years          50         265
##  9:    Estimate!!Total:!!Female:!!30 to 34 years          48         192
## 10:    Estimate!!Total:!!Female:!!35 to 39 years          68         132
## 11:    Estimate!!Total:!!Female:!!40 to 44 years          24         129
## 12:    Estimate!!Total:!!Female:!!45 to 49 years         134          40
## 13:      Estimate!!Total:!!Female:!!5 to 9 years          71         161
## 14:    Estimate!!Total:!!Female:!!50 to 54 years          53         106
## 15:    Estimate!!Total:!!Female:!!55 to 59 years          79         199
## 16:   Estimate!!Total:!!Female:!!60 and 61 years          19          68
## 17:    Estimate!!Total:!!Female:!!62 to 64 years          30          91
## 18:   Estimate!!Total:!!Female:!!65 and 66 years           9          27
## 19:    Estimate!!Total:!!Female:!!67 to 69 years          30         120
## 20:    Estimate!!Total:!!Female:!!70 to 74 years          21          64
## 21:    Estimate!!Total:!!Female:!!75 to 79 years           6          90
## 22:    Estimate!!Total:!!Female:!!80 to 84 years          19          30
## 23: Estimate!!Total:!!Female:!!85 years and over          24           5
## 24:     Estimate!!Total:!!Female:!!Under 5 years          92         159
```

```
##                                                   label 48201312200 48201310300
##      48201411900
##   1:        1559
##   2:          41
##   3:          20
##   4:           0
##   5:          18
##   6:           0
##   7:          53
##   8:          94
##   9:         115
##  10:          44
##  11:         139
##  12:         160
##  13:          39
##  14:         115
##  15:         158
##  16:          45
##  17:          67
##  18:          59
##  19:          71
##  20:          91
##  21:          85
##  22:          17
##  23:          19
##  24:         109
##      48201411900
```

```
BF_SAR_2017 <- SAR_2017[str_detect(label,"Female")&str_detect(concept,"BLACK")]
BF3_SAR_2017 <- BF_SAR_2017[order(label),
                  list(`label`,`48201312200`,`48201310300`,`48201411900`)]
BF3_SAR_2017
```

```
##                                                   label 48201312200 48201310300
##   1:                      Estimate!!Total:!!Female:         814         130
##   2:    Estimate!!Total:!!Female:!!10 to 14 years           1           0
##   3:    Estimate!!Total:!!Female:!!15 to 17 years          11           0
##   4:   Estimate!!Total:!!Female:!!18 and 19 years          24           0
##   5:    Estimate!!Total:!!Female:!!20 to 24 years          84           0
##   6:    Estimate!!Total:!!Female:!!25 to 29 years          50          58
##   7:    Estimate!!Total:!!Female:!!30 to 34 years          48          25
##   8:    Estimate!!Total:!!Female:!!35 to 44 years          83           9
##   9:    Estimate!!Total:!!Female:!!45 to 54 years         164           0
##  10:      Estimate!!Total:!!Female:!!5 to 9 years          71           0
##  11:    Estimate!!Total:!!Female:!!55 to 64 years         121          14
##  12:    Estimate!!Total:!!Female:!!65 to 74 years          51           0
##  13:    Estimate!!Total:!!Female:!!75 to 84 years          18           0
```

```
## 14: Estimate!!Total:!!Female:!!85 years and over          24           0
## 15:     Estimate!!Total:!!Female:!!Under 5 years          64          24
##     48201411900
##  1:           0
##  2:           0
##  3:           0
##  4:           0
##  5:           0
##  6:           0
##  7:           0
##  8:           0
##  9:           0
## 10:           0
## 11:           0
## 12:           0
## 13:           0
## 14:           0
## 15:           0
```

```r
err_sex_by_age_race_data_from_census_17 <-
      censusData_byGroupName(censusdir, vintage, state, censuskey,
                             groupname = "B01001",county_num = "201",
                             block="tract",api_type="acs/acs5",path_suff="err.csv")
```

```
## [1] "found folder ~/Downloads/UH_OneDrive/OneDrive - University Of Houston/Social Network Hy
## [1] "Reading file from ~/Downloads/UH_OneDrive/OneDrive - University Of Houston/Social Netwo
```

```r
errSAR_2017 <- as.data.table(err_sex_by_age_race_data_from_census_17)
errF_SAR_2017 <- errSAR_2017[str_detect(label,"Female")&!str_detect(concept,"O")]
errF3_SAR_2017 <- errF_SAR_2017[order(label),
                  list(`label`,`48201312200`,`48201310300`,`48201411900`)]
errF3_SAR_2017
```

```
##                                         label 48201312200 48201310300
##  1:                   Margin of Error:!!Female:         223         257
##  2:   Margin of Error:!!Female:!!10 to 14 years           4          76
##  3:   Margin of Error:!!Female:!!15 to 17 years          20          55
##  4:  Margin of Error:!!Female:!!18 and 19 years          39          33
##  5:         Margin of Error:!!Female:!!20 years          51          14
##  6:         Margin of Error:!!Female:!!21 years          16          31
##  7:   Margin of Error:!!Female:!!22 to 24 years          44          76
##  8:   Margin of Error:!!Female:!!25 to 29 years          42         118
##  9:   Margin of Error:!!Female:!!30 to 34 years          50          70
## 10:   Margin of Error:!!Female:!!35 to 39 years          59          72
## 11:   Margin of Error:!!Female:!!40 to 44 years          33          93
## 12:   Margin of Error:!!Female:!!45 to 49 years          64          40
```

```
## 13:      Margin of Error:!!Female:!!5 to 9 years          63         107
## 14:    Margin of Error:!!Female:!!50 to 54 years          41          69
## 15:    Margin of Error:!!Female:!!55 to 59 years          52          95
## 16:   Margin of Error:!!Female:!!60 and 61 years          23          49
## 17:    Margin of Error:!!Female:!!62 to 64 years          33          60
## 18:   Margin of Error:!!Female:!!65 and 66 years          10          32
## 19:    Margin of Error:!!Female:!!67 to 69 years          21          69
## 20:    Margin of Error:!!Female:!!70 to 74 years          22          47
## 21:    Margin of Error:!!Female:!!75 to 79 years          12          82
## 22:    Margin of Error:!!Female:!!80 to 84 years          23          28
## 23: Margin of Error:!!Female:!!85 years and over          25          17
## 24:     Margin of Error:!!Female:!!Under 5 years          72          82
##                                             label 48201312200 48201310300
##     48201411900
##  1:         274
##  2:          38
##  3:          30
##  4:          14
##  5:          27
##  6:          14
##  7:          58
##  8:          61
##  9:          71
## 10:          34
## 11:         134
## 12:          82
## 13:          48
## 14:         109
## 15:          79
## 16:          48
## 17:          57
## 18:          56
## 19:          58
## 20:         105
## 21:          52
## 22:          27
## 23:          29
## 24:          66
##     48201411900
```

```
errBF_SAR_2017 <- errSAR_2017[str_detect(label,"Female")&str_detect(concept,"BLACK")]
errBF3_SAR_2017 <- errBF_SAR_2017[order(label),
                      list(`label`,`48201312200`,`48201310300`,`48201411900`)]
errBF3_SAR_2017
```

```
##                                             label 48201312200 48201310300
##  1:                    Margin of Error:!!Female:         222          80
```

```
## 2:    Margin of Error:!!Female:!!10 to 14 years          4          14
## 3:    Margin of Error:!!Female:!!15 to 17 years         20          14
## 4:   Margin of Error:!!Female:!!18 and 19 years         39          14
## 5:    Margin of Error:!!Female:!!20 to 24 years         67          14
## 6:    Margin of Error:!!Female:!!25 to 29 years         42          58
## 7:    Margin of Error:!!Female:!!30 to 34 years         50          29
## 8:    Margin of Error:!!Female:!!35 to 44 years         66          16
## 9:    Margin of Error:!!Female:!!45 to 54 years         70          14
## 10:     Margin of Error:!!Female:!!5 to 9 years         63          14
## 11:    Margin of Error:!!Female:!!55 to 64 years         62          20
## 12:    Margin of Error:!!Female:!!65 to 74 years         29          14
## 13:    Margin of Error:!!Female:!!75 to 84 years         20          14
## 14: Margin of Error:!!Female:!!85 years and over         25          14
## 15:     Margin of Error:!!Female:!!Under 5 years         77          26
##     48201411900
##  1:          14
##  2:          14
##  3:          14
##  4:          14
##  5:          14
##  6:          14
##  7:          14
##  8:          14
##  9:          14
## 10:          14
## 11:          14
## 12:          14
## 13:          14
## 14:          14
## 15:          14
```

```
sex_by_age_race_data_from_census_18 <-
        censusData_byGroupName(censusdir, vintage="2018", state, censuskey,
                               groupname = "B01001",county_num = "201",
                               block="tract",api_type="acs/acs5",path_suff="est.csv")
```

```
## [1] "found folder ~/Downloads/UH_OneDrive/OneDrive - University Of Houston/Social Network Hy
## [1] "Reading file from ~/Downloads/UH_OneDrive/OneDrive - University Of Houston/Social Netwo
```

```
SAR_2018 <- as.data.table(sex_by_age_race_data_from_census_18)
#all the designations have o in them except totals for all races
F_SAR_2018 <- SAR_2018[str_detect(label,"Female")&!str_detect(concept,"O")]
F3_SAR_2018 <- F_SAR_2018[order(label),
                list(`label`,`48201312200`,`48201310300`,`48201411900`)]
F3_SAR_2018
```

```
##                                          label 48201312200 48201310300
```

```
## 1:                   Estimate!!Total!!Female         991        2176
## 2:    Estimate!!Total!!Female!!10 to 14 years          32          82
## 3:    Estimate!!Total!!Female!!15 to 17 years          25         109
## 4:   Estimate!!Total!!Female!!18 and 19 years          19          45
## 5:          Estimate!!Total!!Female!!20 years           9           0
## 6:          Estimate!!Total!!Female!!21 years           9          20
## 7:    Estimate!!Total!!Female!!22 to 24 years          32         123
## 8:    Estimate!!Total!!Female!!25 to 29 years          23         182
## 9:    Estimate!!Total!!Female!!30 to 34 years          80         164
## 10:    Estimate!!Total!!Female!!35 to 39 years        118         244
## 11:    Estimate!!Total!!Female!!40 to 44 years         24         163
## 12:    Estimate!!Total!!Female!!45 to 49 years        117          63
## 13:      Estimate!!Total!!Female!!5 to 9 years        109         166
## 14:    Estimate!!Total!!Female!!50 to 54 years         64         140
## 15:    Estimate!!Total!!Female!!55 to 59 years         68         134
## 16:   Estimate!!Total!!Female!!60 and 61 years         18          44
## 17:    Estimate!!Total!!Female!!62 to 64 years         40          74
## 18:   Estimate!!Total!!Female!!65 and 66 years         12          16
## 19:    Estimate!!Total!!Female!!67 to 69 years         42          76
## 20:    Estimate!!Total!!Female!!70 to 74 years         25          80
## 21:    Estimate!!Total!!Female!!75 to 79 years         10          47
## 22:    Estimate!!Total!!Female!!80 to 84 years         18          36
## 23: Estimate!!Total!!Female!!85 years and over         11          23
## 24:     Estimate!!Total!!Female!!Under 5 years         86         145
##                                          label 48201312200 48201310300
##     48201411900
## 1:        1440
## 2:          39
## 3:           0
## 4:           0
## 5:           0
## 6:           0
## 7:          71
## 8:          61
## 9:         154
## 10:          34
## 11:         172
## 12:         105
## 13:          13
## 14:          90
## 15:         172
## 16:         108
## 17:          51
## 18:          36
## 19:          74
## 20:          79
## 21:          58
## 22:          15
```

```
## 23:          35
## 24:          73
##     48201411900
```

```
BF_SAR_2018 <- SAR_2018[str_detect(label,"Female")&str_detect(concept,"BLACK")]
BF3_SAR_2018 <- BF_SAR_2018[order(label),
                  list(`label`,`48201312200`,`48201310300`,`48201411900`)]
BF3_SAR_2018
```

```
##                                         label 48201312200 48201310300
##  1:                     Estimate!!Total!!Female         786         133
##  2:    Estimate!!Total!!Female!!10 to 14 years           6           0
##  3:    Estimate!!Total!!Female!!15 to 17 years          11           0
##  4:   Estimate!!Total!!Female!!18 and 19 years          19           0
##  5:    Estimate!!Total!!Female!!20 to 24 years          41           0
##  6:    Estimate!!Total!!Female!!25 to 29 years          23          49
##  7:    Estimate!!Total!!Female!!30 to 34 years          80          22
##  8:    Estimate!!Total!!Female!!35 to 44 years          88          20
##  9:    Estimate!!Total!!Female!!45 to 54 years         160           0
## 10:      Estimate!!Total!!Female!!5 to 9 years          84           0
## 11:    Estimate!!Total!!Female!!55 to 64 years         120           8
## 12:    Estimate!!Total!!Female!!65 to 74 years          75           0
## 13:    Estimate!!Total!!Female!!75 to 84 years          22           0
## 14: Estimate!!Total!!Female!!85 years and over          11           0
## 15:     Estimate!!Total!!Female!!Under 5 years          46          34
##     48201411900
##  1:          54
##  2:           0
##  3:           0
##  4:           0
##  5:           0
##  6:           0
##  7:           0
##  8:           0
##  9:           0
## 10:           0
## 11:          54
## 12:           0
## 13:           0
## 14:           0
## 15:           0
```

Look at how errors correlate across years

```
err_sex_by_age_race_data_from_census_18 <-
        censusData_byGroupName(censusdir, vintage="2018", state, censuskey,
```

```r
                                    groupname = "B01001",county_num = "201",
                                    block="tract",api_type="acs/acs5",path_suff="err.csv")
```

```
## [1] "found folder ~/Downloads/UH_OneDrive/OneDrive - University Of Houston/Social Network Hy
## [1] "Reading file from ~/Downloads/UH_OneDrive/OneDrive - University Of Houston/Social Netwo
```

```r
errSAR_2018 <- as.data.table(err_sex_by_age_race_data_from_census_18)
errF_SAR_2018 <- errSAR_2018[str_detect(label,"Female")&!str_detect(concept,"O")]
errF3_SAR_2018 <- errF_SAR_2018[order(label),
                    list(`label`,`48201312200`,`48201310300`,`48201411900`)]
errF3_SAR_2018
```

```
##                                         label 48201312200 48201310300
##  1:                      Margin of Error!!Female         260         290
##  2:    Margin of Error!!Female!!10 to 14 years          27          60
##  3:    Margin of Error!!Female!!15 to 17 years          27          90
##  4:   Margin of Error!!Female!!18 and 19 years          29          57
##  5:          Margin of Error!!Female!!20 years          13          13
##  6:          Margin of Error!!Female!!21 years          13          31
##  7:    Margin of Error!!Female!!22 to 24 years          40          82
##  8:    Margin of Error!!Female!!25 to 29 years          24          84
##  9:    Margin of Error!!Female!!30 to 34 years          69          67
## 10:    Margin of Error!!Female!!35 to 39 years          55          87
## 11:    Margin of Error!!Female!!40 to 44 years          31         101
## 12:    Margin of Error!!Female!!45 to 49 years          58          50
## 13:      Margin of Error!!Female!!5 to 9 years          57         125
## 14:    Margin of Error!!Female!!50 to 54 years          49          81
## 15:    Margin of Error!!Female!!55 to 59 years          47          74
## 16:   Margin of Error!!Female!!60 and 61 years          23          37
## 17:    Margin of Error!!Female!!62 to 64 years          40          47
## 18:   Margin of Error!!Female!!65 and 66 years          12          25
## 19:    Margin of Error!!Female!!67 to 69 years          26          56
## 20:    Margin of Error!!Female!!70 to 74 years          26          55
## 21:    Margin of Error!!Female!!75 to 79 years          12          52
## 22:    Margin of Error!!Female!!80 to 84 years          23          30
## 23: Margin of Error!!Female!!85 years and over          12          30
## 24:     Margin of Error!!Female!!Under 5 years          70          83
##                                         label 48201312200 48201310300
##     48201411900
##  1:         222
##  2:          35
##  3:          13
##  4:          13
##  5:          13
##  6:          13
##  7:          59
```

```
## 8:           57
## 9:           74
## 10:          33
## 11:         139
## 12:          85
## 13:          23
## 14:          95
## 15:          69
## 16:          97
## 17:          51
## 18:          39
## 19:          55
## 20:          91
## 21:          44
## 22:          25
## 23:          40
## 24:          45
##     48201411900
```

Look at how the margin of error numbers compare between 2017-2018

```
errF3_SAR_2018[,2:4]-errF3_SAR_2017[,2:4]
```

```
##     48201312200 48201310300 48201411900
##  1:          37          33         -52
##  2:          23         -16          -3
##  3:           7          35         -17
##  4:         -10          24          -1
##  5:         -38          -1         -14
##  6:          -3           0          -1
##  7:          -4           6           1
##  8:         -18         -34          -4
##  9:          19          -3           3
## 10:          -4          15          -1
## 11:          -2           8           5
## 12:          -6          10           3
## 13:          -6          18         -25
## 14:           8          12         -14
## 15:          -5         -21         -10
## 16:           0         -12          49
## 17:           7         -13          -6
## 18:           2          -7         -17
## 19:           5         -13          -3
## 20:           4           8         -14
## 21:           0         -30          -8
## 22:           0           2          -2
## 23:         -13          13          11
```

```
## 24:          -2           1         -21
##     48201312200 48201310300 48201411900
```

```
errBF_SAR_2018 <- errSAR_2018[str_detect(label,
                      "Female")&str_detect(concept,"BLACK")]
errBF3_SAR_2018 <- errBF_SAR_2018[order(label),
                      list(`label`,`48201312200`,`48201310300`,`48201411900`)]
errBF3_SAR_2018
```

```
##                                          label 48201312200 48201310300
##  1:                     Margin of Error!!Female         222          72
##  2:    Margin of Error!!Female!!10 to 14 years           9          13
##  3:    Margin of Error!!Female!!15 to 17 years          15          13
##  4:   Margin of Error!!Female!!18 and 19 years          29          13
##  5:    Margin of Error!!Female!!20 to 24 years          43          13
##  6:    Margin of Error!!Female!!25 to 29 years          24          58
##  7:    Margin of Error!!Female!!30 to 34 years          69          25
##  8:    Margin of Error!!Female!!35 to 44 years          61          22
##  9:    Margin of Error!!Female!!45 to 54 years          72          13
## 10:      Margin of Error!!Female!!5 to 9 years          58          13
## 11:    Margin of Error!!Female!!55 to 64 years          65          14
## 12:    Margin of Error!!Female!!65 to 74 years          36          13
## 13:    Margin of Error!!Female!!75 to 84 years          23          13
## 14: Margin of Error!!Female!!85 years and over          12          13
## 15:     Margin of Error!!Female!!Under 5 years          57          31
##     48201411900
##  1:          85
##  2:          13
##  3:          13
##  4:          13
##  5:          13
##  6:          13
##  7:          13
##  8:          13
##  9:          13
## 10:          13
## 11:          85
## 12:          13
## 13:          13
## 14:          13
## 15:          13
```

Look at how the margin of error numbers compare between 2017-2018

```
errBF3_SAR_2018[,2:4]-errBF3_SAR_2017[,2:4]
```

```
##     48201312200 48201310300 48201411900
```

```
##  1:           0           -8          71
##  2:           5           -1          -1
##  3:          -5           -1          -1
##  4:         -10           -1          -1
##  5:         -24           -1          -1
##  6:         -18            0          -1
##  7:          19           -4          -1
##  8:          -5            6          -1
##  9:           2           -1          -1
## 10:          -5           -1          -1
## 11:           3           -6          71
## 12:           7           -1          -1
## 13:           3           -1          -1
## 14:         -13           -1          -1
## 15:         -20            5          -1
```

For females and using the Census' guide for calculating standard error from the given margin of error: Standard Error = Margin of Error / Z, where Z = 1.645 for census products after 2005.

```r
errF3_SAR_2018[,2:4]/1.645
```

```
##     48201312200 48201310300 48201411900
##  1:  158.054711  176.291793  134.954407
##  2:   16.413374   36.474164   21.276596
##  3:   16.413374   54.711246    7.902736
##  4:   17.629179   34.650456    7.902736
##  5:    7.902736    7.902736    7.902736
##  6:    7.902736   18.844985    7.902736
##  7:   24.316109   49.848024   35.866261
##  8:   14.589666   51.063830   34.650456
##  9:   41.945289   40.729483   44.984802
## 10:   33.434650   52.887538   20.060790
## 11:   18.844985   61.398176   84.498480
## 12:   35.258359   30.395137   51.671733
## 13:   34.650456   75.987842   13.981763
## 14:   29.787234   49.240122   57.750760
## 15:   28.571429   44.984802   41.945289
## 16:   13.981763   22.492401   58.966565
## 17:   24.316109   28.571429   31.003040
## 18:    7.294833   15.197568   23.708207
## 19:   15.805471   34.042553   33.434650
## 20:   15.805471   33.434650   55.319149
## 21:    7.294833   31.610942   26.747720
## 22:   13.981763   18.237082   15.197568
## 23:    7.294833   18.237082   24.316109
## 24:   42.553191   50.455927   27.355623
##     48201312200 48201310300 48201411900
```

Standard error Black female

```
errBF3_SAR_2018[,2:4]/1.645
```

```
##      48201312200 48201310300 48201411900
##   1:  134.954407   43.768997   51.671733
##   2:    5.471125    7.902736    7.902736
##   3:    9.118541    7.902736    7.902736
##   4:   17.629179    7.902736    7.902736
##   5:   26.139818    7.902736    7.902736
##   6:   14.589666   35.258359    7.902736
##   7:   41.945289   15.197568    7.902736
##   8:   37.082067   13.373860    7.902736
##   9:   43.768997    7.902736    7.902736
##  10:   35.258359    7.902736    7.902736
##  11:   39.513678    8.510638   51.671733
##  12:   21.884498    7.902736    7.902736
##  13:   13.981763    7.902736    7.902736
##  14:    7.294833    7.902736    7.902736
##  15:   34.650456   18.844985    7.902736
```

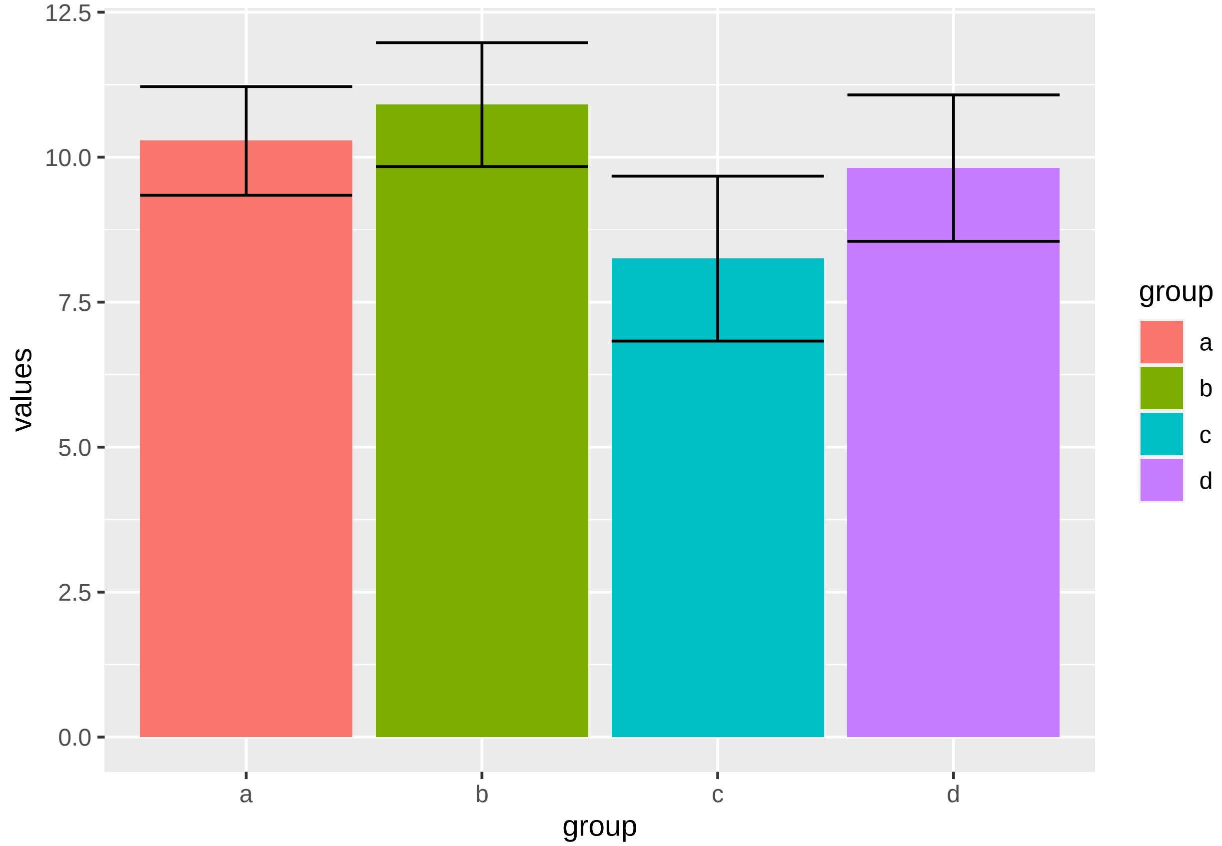
And then the estimated value minus the standard error for 2018

```
BF3_SAR_2018[,2:4]-(errBF3_SAR_2018[,2:4]/1.645)
```

```
##      48201312200 48201310300 48201411900
##   1: 651.0455927  89.2310030    2.328267
##   2:   0.5288754  -7.9027356   -7.902736
##   3:   1.8814590  -7.9027356   -7.902736
##   4:   1.3708207  -7.9027356   -7.902736
##   5:  14.8601824  -7.9027356   -7.902736
##   6:   8.4103343  13.7416413   -7.902736
##   7:  38.0547112   6.8024316   -7.902736
##   8:  50.9179331   6.6261398   -7.902736
##   9: 116.2310030  -7.9027356   -7.902736
##  10:  48.7416413  -7.9027356   -7.902736
##  11:  80.4863222  -0.5106383    2.328267
##  12:  53.1155015  -7.9027356   -7.902736
##  13:   8.0182371  -7.9027356   -7.902736
##  14:   3.7051672  -7.9027356   -7.902736
##  15:  11.3495441  15.1550152   -7.902736
```

If we assumed that there was a sort of expected variation

```r
#from https://statisticsglobe.com/add-standard-error-bars-barchart-r - with diff numbers
library(ggplot2)
df_example <- data.frame(values = rnorm(100,10,7),group = letters[1:4])
ggplot(df_example, aes(values, group, fill = group)) +
  coord_flip() +
  stat_summary(geom = "bar", fun = mean, position = "dodge") +
  stat_summary(geom = "errorbar", fun.data = mean_se, position = "dodge")
```



If we do the setup for calculating standard error and map it

```r
#left_join -example_data would be ... talk to Ioannis about what to show - have to just pick o
```

Line up each of the years as stacked by age, starting with 2010, and then go to next year

```r
#left_join()
```

Clearly, this is a deep problem for using the American Community Survey for small area estimation - or more generally, for the modeling that health professionals (and community members and students) would like to do to understand what health interventions are more likely to be effective. For example, new data tools are being developed and bulk data downloads made available for small area estimation of health outcomes. This included some help on ways to create health rankings within cities from the data modeled by the 500 Cities Project (https://www.cdc.gov/places/about/500-

cities-2016-2019/index.html), later replaced by the Places project (https://www.cdc.gov/places/index.html). At first glance, this is what Sam City was also supposed to give us, but the published approaches had not addressed any of our concerns, either the philosophical ones we will examine in the next part or the simpler ones about continuity that we just saw in the numbers assigned to the tracts. The proliferation of other sites that made the same data available in slightly different forms, often for homebuyers and not policy-makers (https://www.cityhealthdashboard.com/, http://www.city-data.com/, https://www.neighborhoodscout.com/, https://www.trulia.com/neighborhoods/, https://www.neighborhoodatlas.medicine.wisc.edu/), added to our confusion about messaging and put into doubt the utility of Sam City even for limited pedagogical uses. The official caveats on use of the ACS suggest complex statistical tests on each level, and take no responsibility for bad uses. A buried caveat about confidence intervals and margins of error will not dissuade someone from creating an automatic map that makes it look like diabetes or maternal health has changed in a particular neighborhood, when in fact indiscriminately (and unknowably) large effects within the analysis would be an artifact of the choices that were made in creating that map - and specifically, the mapping of the larger numbers at aggregated levels of analysis onto the smaller areas that constitute our daily places of engagement.

To just point out one of the most obvious choices, in those mappings, the census wanted to preserve the statistical structure at certain levels and was willing to sacrifice other structures in order to keep that broad horizon of being able to justify each step in terms of a representation of statistical likelihood relative to any particular combination instead of seeing the problem as how to optimize distribution among potential categories (either real or conceptual spaces). We learned this at great expense - and very great frustration for a gifted student who spent many hours trying to make it work in an early version of Sam City. She had been asked by our faculty team to create the pedagogical tool by calculating the percentage chance for any individual to be found in the next category of interest, and then to distribute them by that likelihood. She would try to create ever more complicated examples, but always ran into insurmountable walls as the pieces refused to fall into place. We later stepped back and looked at the problem again. We saw that regardless of our view of the ultimate horizon of truth or falsity, we were dealing with a certain type of game, where the problem was to put people into spaces (conceptual and real) that recaptured the original dispensation of people in those spaces (which was, itself, a bit of a game).

We are inspired here by certain quite technical innovation in mathematics (cf. https://arxiv.org/abs/1703.03007 for an overview on homotopy type theory and conceptual spaces) and in statistics, especially as related to language (cf. T-D Bradley, https://arxiv.org/abs/2004.05631, and https://arxiv.org/abs/2106.07890). Lawvere's own intro to math is also very much about spaces. We also hope to have some concrete answers to problems in small area estimation. https://datascience.codata.org/articles/10.5334/dsj-2018-008/ could be a starting point for that.

Perhaps example of Hispanic ethnicity/race and how they have to add up?

```
#could also embed these in the next steps...
```

So how do we fix this? Next part is "Making Sam"