

Social Service Agency

Andy Grogan-Kaylor

March 30, 2021

Contents

0.1 Background	1
0.2 Clients	1
0.3 Donors	3

0.1 Background

You are working for a social service agency that keeps track of its clients and donors using Excel. While the agency has kept track of these two groups of individuals, the agency has done little to evaluate their work, and to assess ways in which they might do better.

The agency would like you to provide them with some data visualization, and they will provide you with a file of **clients** (in 2 formats)¹, and a file of **donors**.

0.2 Clients

Here is what the data file of clients looks like.

NB that the mental health instrument is normed such that **100** is considered to be an *average* or *usual* mental health score. Higher scores indicate higher psychological wellbeing.

¹This is an **advanced** topic, but it is worth thinking the idea that sometimes data visualization problems are really problems of how the underlying data is structured. Initially, the data called **clients** is easier to use, but for some purposes **clients_longform** is easier to work with. The two files contain the same *information*. Are there circumstances in which one would be easier to work with than the other?

Table 1: Table continues below

ID	age	gender	race_ethnicity	family_income	program
2844	34	Other Identity	African American	62246	Program C
1780	35	Other Identity	African American	34123	Program C
1059	18	Male	Asian American	61076	Program B
3729	25	Female	African American	59285	Program A
3151	20	Male	Asian American	40630	Program A
1817	20	Other Identity	African American	54774	Program C

mental_health_T1	mental_health_T2	latitude	longitude
90.65	91.59	42.16	-83.75
81.67	89.52	42.26	-83.86
95.47	102.6	42.36	-83.72
87.82	94.49	42.32	-83.82
91.38	93.56	42.2	-83.76
85.61	97.59	42.29	-83.62

The agency also had an intern from a school of social work who did some preliminary data visualization.

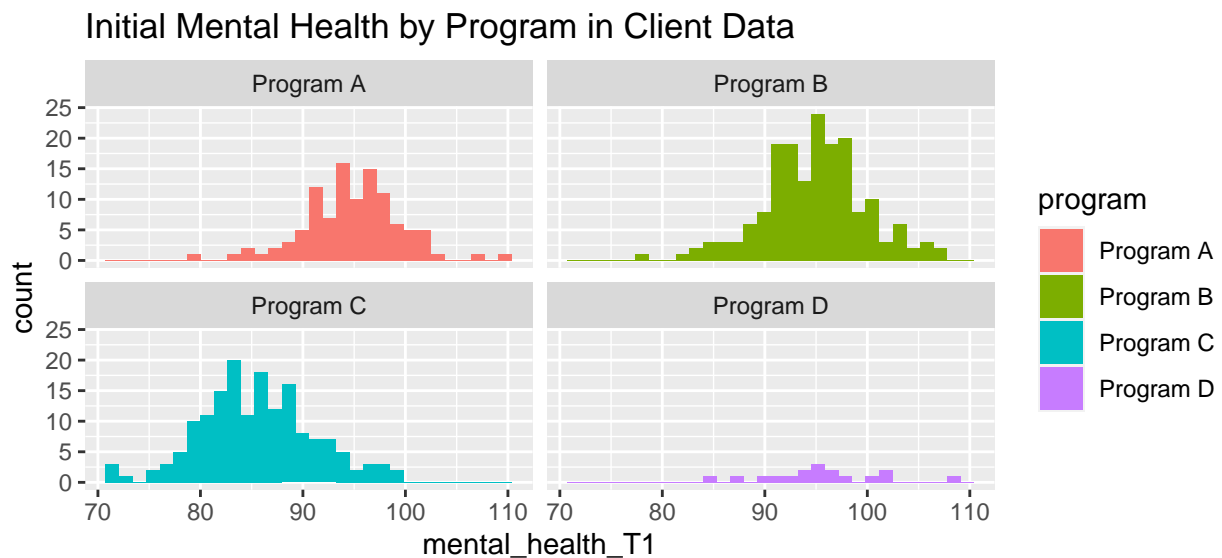


Figure 1: Preliminary Data Visualization of Clients

Agency staff would like to investigate questions of the following nature.

- What are demographic characteristics of clients?

- Do any demographic characteristics of clients have an association with mental health or program use?
- Is there any additional information that the agency should be collecting about clients?
- Are any of the programs associated with better mental health at Time 2?

The Director of the agency says:

“I’ve heard you can investigate things like this with `ggplot()`, maybe using something like `geom_point()`, `geom_smooth()`, and `facet_wrap(~)`.”²

0.3 Donors

The agency also has a file of donors which looks something like this:

ID	age	gender	neighborhood	annual_donation
rllTxfM	54.3	Male	Neighborhood A	60
KUbokOQ	37.55	Male	Neighborhood A	98
ksGEKpd	35.99	Female	Neighborhood A	256
TokbALD	31.32	Male	Neighborhood A	63
SS6vuYl	53.43	Male	Neighborhood A	381
xDvAUym	35.73	Female	Neighborhood B	1202

The agency would like to know.

- In terms of available information, what are the demographic characteristics of donors?
- Is there any additional information that we should collect about donors?
- Is there any information that predicts which donors will donate larger amounts of money?

The agency did hire a consulting agency to do some preliminary analysis of their donors. After paying the consultant several thousand dollars they received a short report and the attached graphic.

The Director says:

“While the attached graphic may give you, and us, some guidance, we are wondering if you can produce anything that is a little simpler, a little more straightforward, and a little more clear.”

²Yes, the Director really talks like this sometimes.



Figure 2: Preliminary Visualization of Donors