

Report on Survey Weights

Maksim Zubok

April 22, 2024

Contents

Introduction	1
Sample to Population Comparison	1
Weights with Survey package	3
Weights created manually	7

Introduction

To create weights for all surveys, we are working with the 2020 census data, particularly the cross tabbed gender, age, and university education file [here](#).

We do not include information about region of residence, even though we could do it after harmonising census data with the survey.

Table 1: Population Frame: Census 2020

gender	age_group	university_education	Freq
Man	18-24	BA+	463546
Man	18-24	BA-	5164261
Man	25-34	BA+	2707302
Man	25-34	BA-	7287296
Man	35-44	BA+	3136411

Sample to Population Comparison

To compare our samples to the population, we do two comparisons. First, we compare them to a nationally representative Levada omnibus survey. Second, we compare them to the 2020 Census.

The Table 2 compares the demographic composition of Qualtrics samples collected in August, February, and March to nationally representative Levada Omnibus survey.

Table 2: Comparison of Category Shares by Variable

Variable	Values	Levada	Q. March	Q. February	Q. August
Age	18-24	0.08	0.09	0.09	0.07
Age	25-34	0.17	0.21	0.21	0.22
Age	35-44	0.23	0.18	0.18	0.28
Age	45-54	0.15	0.17	0.17	0.21
Age	55-64	0.18	0.20	0.25	0.10
Age	65+	0.19	0.15	0.10	0.12
Gender	Man	0.44	0.45	0.45	0.49
Gender	Woman	0.56	0.55	0.55	0.51
Education	BA+	0.28	0.54	0.54	0.57
Education	BA-	0.72	0.46	0.46	0.43

The Figure 1 depicts strata shares compared to the 2020 Census.

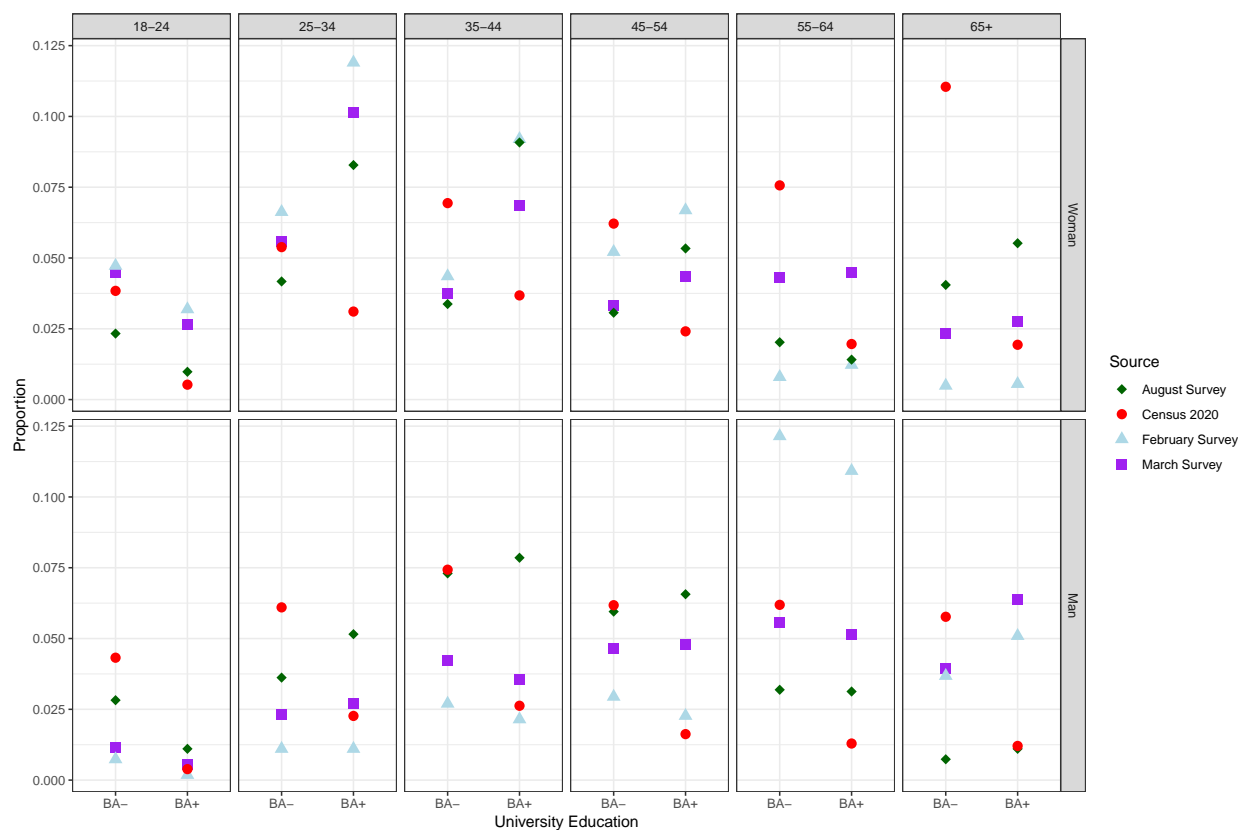


Figure 1: Demographic Comparison on Census 2020

The main observations are:

- The biggest disparities are in women 55-64 and 65+ without education. For instance, the February survey has 0 women in the latter category.
- We have oversampled young women, especially 25-34 years old with university education, and under-sampled men without university education across all age categories except 55- and 65+.

Weights with Survey package

To compute post-stratification weights we rely on the `postStratify` function from the `survey` package. The function adjusts the sampling and replicate weights so that the joint distribution of a set of post-stratifying variables matches the known population joint distribution. **However, the package documentation does not describe how exactly the adjustment is implemented.**

March

Table 3: March Survey PostStratify Weights Summary

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Value	13863.25	24875.41	54076.05	73292.34	107672	347290.3

The table below explores shows strata that were assigned the highest weight.

Table 4: March Survey, Top Five Rows by Weight

age_group	gender	university_education	weight_poststratify
65+	Woman	BA-	347290.3
18-24	Man	BA-	271803.2
25-34	Man	BA-	191770.9
45-54	Woman	BA-	137495.7
35-44	Woman	BA-	135884.6

February

Table 5: February Survey PostStratify Weights Summary

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Value	8670.35	19144.08	37348.11	73292.34	87350.22	1649629

Note: some strata had no observations in the survey (NA on education for some age gender groups). This means we had to ignore them in producing weights.

Table 6: February Survey, Top Five Rows by Unique Weight

age_group	gender	university_education	weight_poststratify
65+	Woman	BA-	1649629.1
55-64	Woman	BA-	695189.1
18-24	Man	BA-	430355.1
25-34	Man	BA-	404849.8
65+	Woman	BA+	256902.0

August

Table 7: August Survey PostStratify Weights Summary

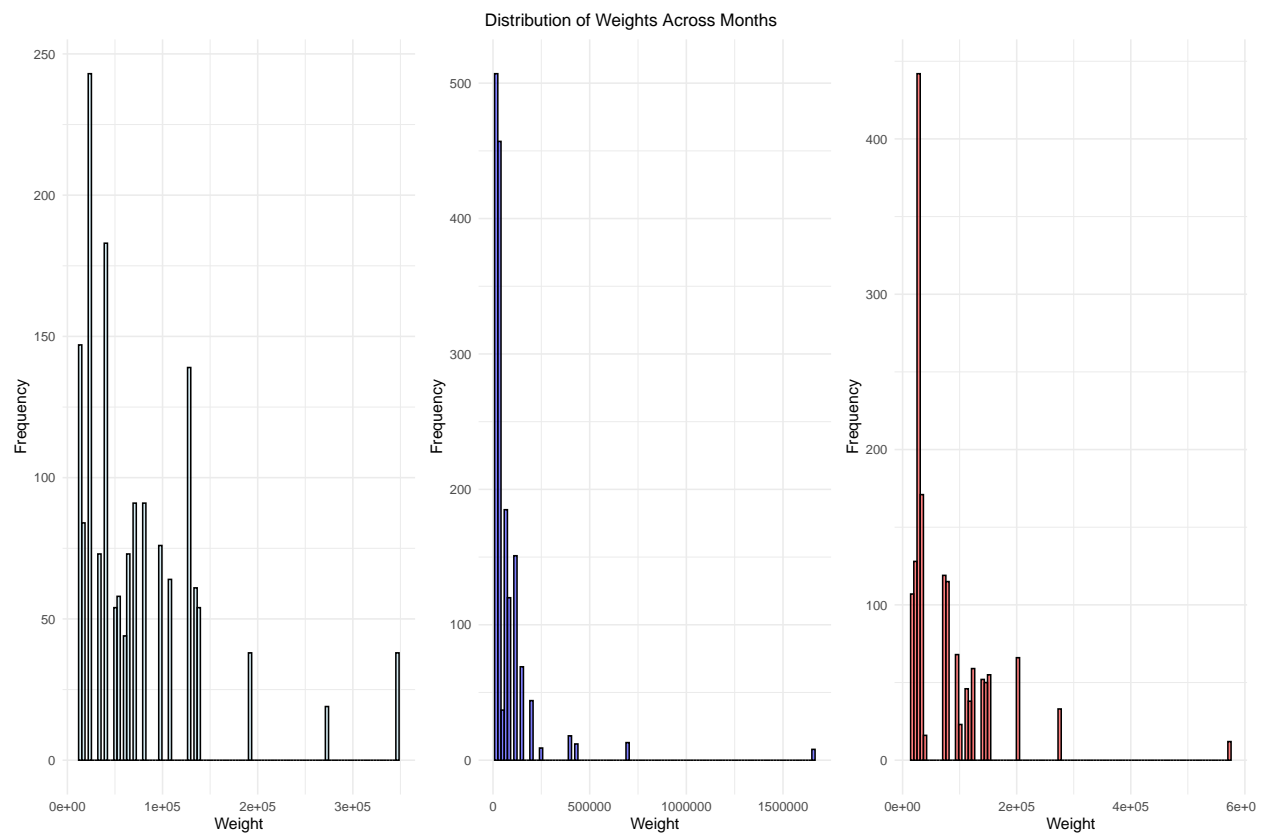
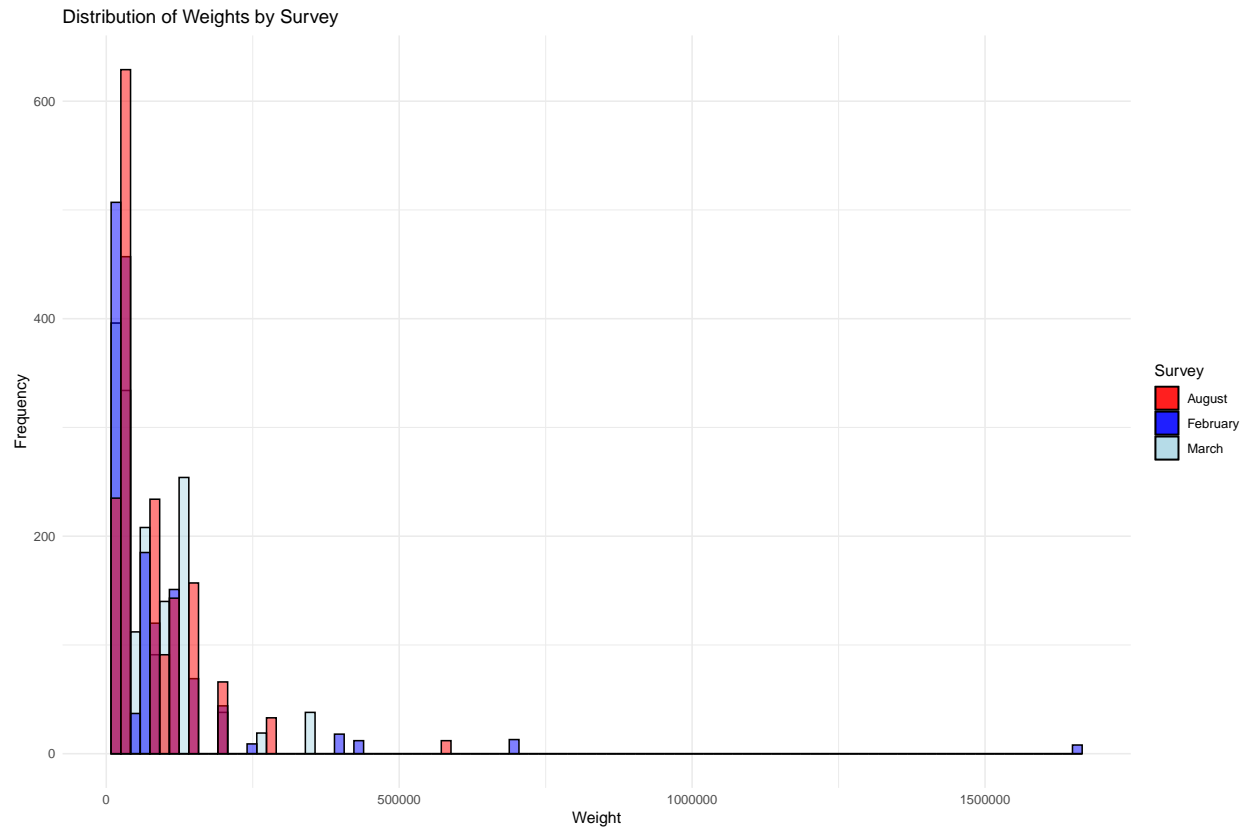
	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Value	18133.48	27510.75	33083.25	74666.57	112266.5	574250.5

For August, we also see that some weights are much larger than others. As you can see in the graphs below, the distribution of weights is similarly skewed and the disparities between the bulk of the distribution and its tails are in the same orders of magnitude. However, the largest weight in Feb survey is three times bigger than the largest weight in Aug survey.

The largest weights in both surveys relate to different population groups.

Table 8: August Survey, Top Five Rows by Weight

age_group	gender	university_education	weight_poststratify
65+	Man	BA-	574250.5
55-64	Woman	BA-	273862.4
65+	Woman	BA-	199955.0
35-44	Woman	BA-	150708.4
45-54	Woman	BA-	148495.4



Weights created manually

To check the plausibility of resulting weights, we create alternative weights based on the population frequencies of the combination of the same strata (Yana's approach). The weights are calculated for each category:

$$\text{weight}_i = \frac{\text{population frequency}_i}{\text{sample frequency}_i}$$

March

Table 9: March Survey Weights Summary

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Value	0.19	0.51	0.85	1.24	1.76	4.74

February

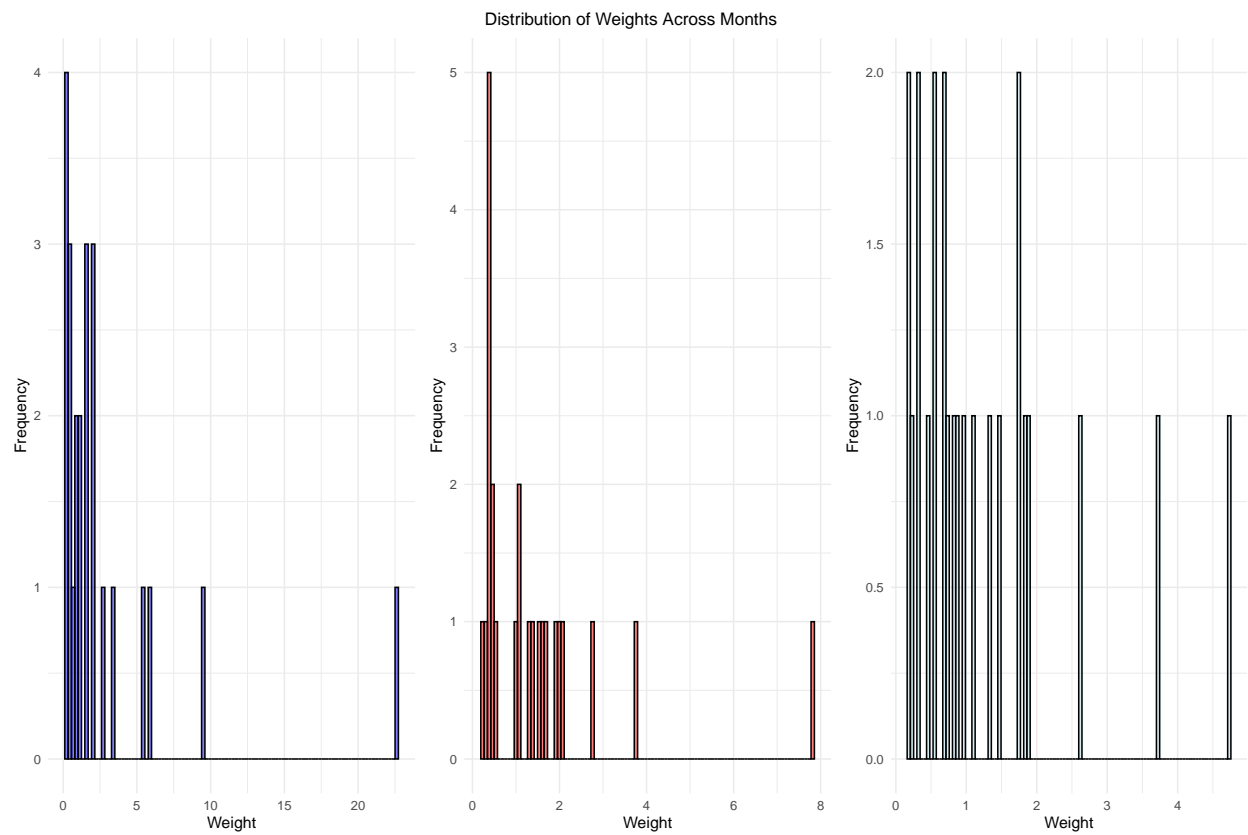
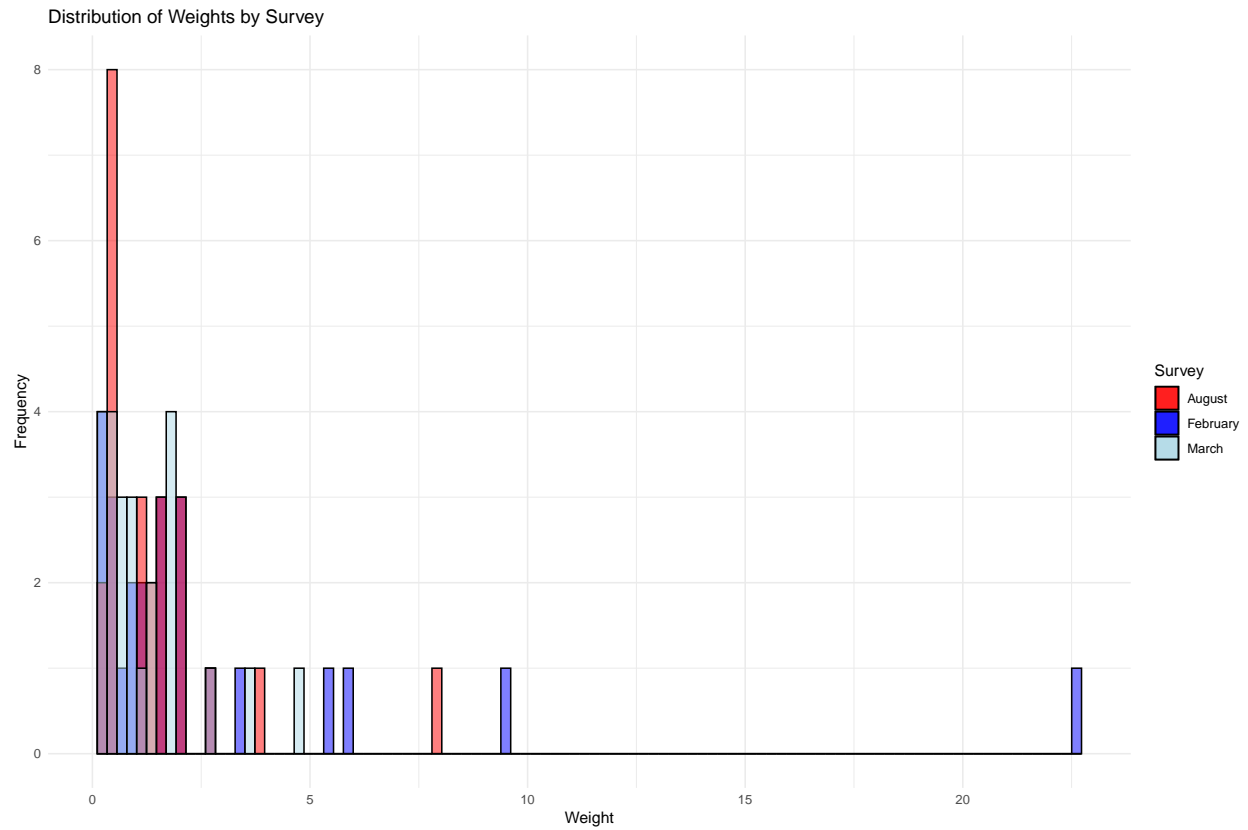
Table 10: February Survey Weights Summary

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Value	0.12	0.48	1.39	2.81	2.27	22.51

August

Table 11: August Survey Weights Summary

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Value	0.25	0.41	1.07	1.46	1.75	7.84



Calculating weights this way gives us the same overall picture. The main difference is in the orders of

magnitude between the median weight and the tail (highest weight).

For **February** survey, the difference (max weight / median weight) with `postStratify` is **44.17**, but with simple manually created weights it is **16.19**. For **August** survey, the difference with `postStratify` is **17.36**, but with simple manually created weights it is **7.33**.