Cassidy Denault
May 2024
Upower Internship - Seattle University

Data Visualizations for Upower

A Brief Overview

Upower is a Seattle based non-profit aimed at providing “trauma-informed coaching using movement and play to empower and educate youth who are underserved”. They provide opportunities for physical activity to youth without such opportunities in their institutions- schools, recovery centers, detention centers, etc. They also educate coaches to be trauma informed, are actively involved in community events, and contribute to research on child development.

As a Data Intern, my main goal was to assist in the processing of their survey data to contribute to the making of their annual report and for internal evaluation of programming. I worked with 2 other student interns, taking point on the project.

Working with likert-like survey and demographic data from youth participants, I was heavily involved with the input, processing, analysis, and communication through visualizations of the data.

Input required: de-identification, codifying, and input of raw data into spreadsheets.

Processing required: manual cleaning/standardization of the raw spreadsheets, writing a script to clean further, making various sub tables and aggregations.

Analysis required: designing methods for measuring impact and change over time for the data, computing descriptive statistics.

Communication required: obtaining an understanding of the target audience, creating visuals (charts and graphs), designing print spreads for a general/donor audience.

Communication and collaboration between Upower mentors and coaches, my fellow data interns, and myself were crucial to my learning and ability to present an accurate data ‘story’.

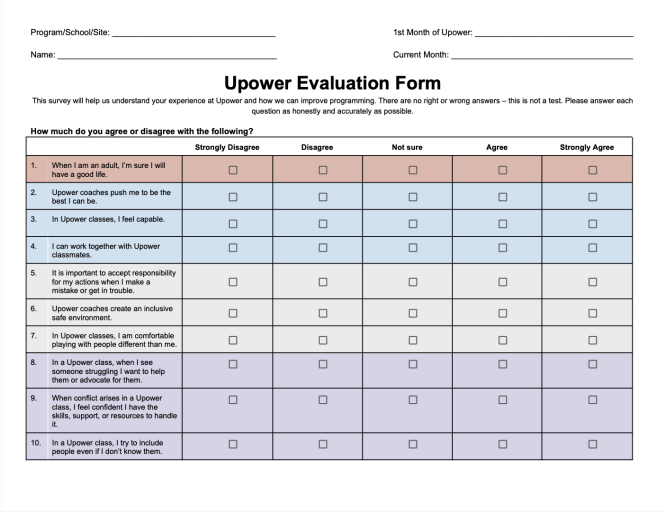
My final contributions are:

- Cleaned data tables
- A pandas python script for cleaning and creating subtables/aggregations
- 3 data driven pages for the annual report
- This report to aid in interpretation of this year’s data and to guide design of next year's surveys

A Brief Project Overview

The data: self reported surveys (beginning, middle, end of year)

Operating on a June-July year cycle, Upower distributes paper surveys to its participants at three points in the year: pre, mid, and post. There are 9 sites (3 which only have post data).



Program/School/Site: _____ 1st Month of Upower: _____

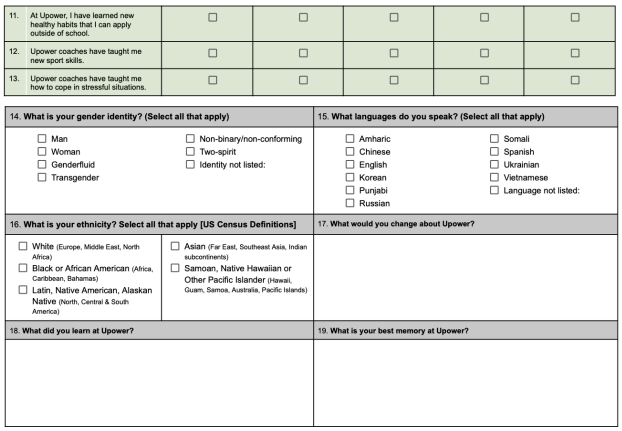
Name: _____ Current Month: _____

Upower Evaluation Form

This survey will help us understand your experience at Upower and how we can improve programming. There are no right or wrong answers – this is not a test. Please answer each question as honestly and accurately as possible.

How much do you agree or disagree with the following?

		Strongly Disagree	Disagree	Not sure	Agree	Strongly Agree
1.	When I am an adult, I'm sure I will have a good life.	☐	☐	☐	☐	☐
2.	Upower coaches push me to be the best I can be.	☐	☐	☐	☐	☐
3.	In Upower classes, I feel capable.	☐	☐	☐	☐	☐
4.	I can work together with Upower classmates.	☐	☐	☐	☐	☐
5.	It is important to accept responsibility for my actions when I make a mistake or get in trouble.	☐	☐	☐	☐	☐
6.	Upower coaches create an inclusive safe environment.	☐	☐	☐	☐	☐
7.	In Upower classes, I am comfortable playing with people different than me.	☐	☐	☐	☐	☐
8.	In a Upower class, when I see someone struggling I want to help them or advocate for them.	☐	☐	☐	☐	☐
9.	When conflict arises in a Upower class, I feel confident I have the skills, support, or resources to handle it.	☐	☐	☐	☐	☐
10.	In a Upower class, I try to include people even if I don't know them.	☐	☐	☐	☐	☐



11.	At Upower, I have learned new healthy habits that I can apply outside of school.	☐	☐	☐	☐	☐
12.	Upower coaches have taught me new sport skills.	☐	☐	☐	☐	☐
13.	Upower coaches have taught me how to cope in stressful situations.	☐	☐	☐	☐	☐

14. What is your gender identity? (Select all that apply)

☐ Man
☐ Woman
☐ Genderfluid
☐ Transgender
☐ Non-binary/non-conforming
☐ Two-spirit
☐ Identity not listed:

15. What languages do you speak? (Select all that apply)

☐ Amharic
☐ Chinese
☐ English
☐ Korean
☐ Punjabi
☐ Russian
☐ Somali
☐ Spanish
☐ Ukrainian
☐ Vietnamese
☐ Language not listed:

16. What is your ethnicity? Select all that apply [US Census Definitions]

☐ White (Europe, Middle East, North Africa)
☐ Black or African American (Africa, Caribbean, Bahamas)
☐ Latin, Native American, Alaskan Native (North, Central & South America)
☐ Asian (Far East, Southeast Asia, Indian subcontinents)
☐ Samoan, Native Hawaiian or Other Pacific Islander (Hawaii, Guam, Samoa, Australia, Pacific Islands)

17. What would you change about Upower?

18. What did you learn at Upower?

19. What is your best memory at Upower?

As you can see from the survey's above, they collect ratings on a 5 point scale in the range of strongly disagree to strongly agree in response to statements related to respondent self perception, Upower programming, etc. It also records the demographic gender identity, language, and ethnicity data (of which respondents can choose multiple).

The questions are grouped into 5 categories:

Confidence (1)
Connection (2, 3, 4)
Character (5, 6, 7)
Caring (8, 9, 10)
Competence (11, 12, 13)

At the start of the project, pre and mid data had already been inputted to spreadsheets. There were sub spreadsheets for each site and for each pre/mid/post surveying. This raw data was not entirely standardized.

Imagine about (9 sites x 3 surveys per year) sub sheets, and inputs for demographics being unstandardized (woman vs women) (samoan vs islander)

Goals for the project

- Evaluate efficacy of Upower programming and effect of Upower on respondents using descriptive statistics.
    - Starting from raw data, do all required input, processing, analyzing, and designing.
- Create visualizations to aid in internal evaluation/reflection *and* to include in an annual report for potential donors.
- Ideate on a standard, maybe automated data intake and processing flow for the coming years

Workflow

A detailed review of the steps I took in this project.

I. Contextualize the project and data
    A. Child development and trauma informed coaching education
    B. Learning about Upower's day to day programming and participants
    C. Generate questions we would like to answer from the data (don't necessarily think of constraints or implementation, brainstorm what we want to know. See Next Steps section for some I posed)

II. Prepare sub spreadsheets for post data input
    A. Carry over demographic data for repeat survey takers

III. Input Post Survey Data
    A. De-identify names and assign codes to surveys.
    B. Manually input scores for questions to spreadsheets. Input demographic data for new participants

IV. Manual Cleaning in sheets
    A. Standardize the ethnicity, language, and gender identity options (some input the same category as different names. "samoan" and "native hawaiian" --> "islander" as they refer to the same option on survey.
    B. Standardize ethnicity, language, and gender identity to the format "choice1,choice2" (no spaces just a comma to separate answers to the same question.
    C. Copy/paste all rows from separate sheets to one sheet. Add columns to specify what site and weather pre, mid, or post

V. Make a python script in google collaboratory to further clean the data and generate subtables and aggregations
    A. Used the manually cleaned sheet
    B. Go from a wide formatted table to a tall formatted table. Instead of each row being for one participant id at one time of survey containing answers for 13 questions, have each question get its own row entry

PARTICIPANT CODE	Good life	coaches push me	feel capable	work together	take responsibility	inclusive safe environment	people different

struggling advocate	conflict support	include people dont know	healthy habits	new sport skills	cope stressful situations	gender identity	languages	ethnicity	site	pre/mid/post

|

| Change to

|

PARTICIPANT CODE	gender identity	languages	ethnicity	site	pre/mid/post	question_id	question_score

In which each question in the colored boxes is assigned a question_id. A participant will have many rows, one for each answered question.

A table with many more rows (but colab is able to handle it) allows for easier filtering based on question_id. This long format is better for data analysis, scripting, and generating graphs and charts- but not as easy for people to understand visually.

C. Make frequency tables and subtables to see things like demographics, amount of participants that increased scores for before and after surveys, etc.
D. See the COLAB SCRIPT for exact processing details

VI. Export clean tables back to sheets to make charts and calculate further descriptive statistics.
A. See the GOOGLE SHEET FILE for exact processing details

VII. Design the annual report data section
A. Choose results relevant for the audience- can't have too many graphs and numbers, but which are most impactful to the message we are trying to send?
B. Choose the types of charts and visuals to best represent findings.
C. Edit the page in Canva, conforming to brand guidelines and principles of design.

Findings & Visualizations

These are the comprehensive results of my analysis of the data. The annual report does not require such a technical analysis of the data, and in fact benefits from overwhelming readers with crowded pages and complicated analysis. However, these can be beneficial internally to Upower as an avenue of feedback from respondents.

As always, consider the limitations of data as a means to tell a complicated story in a specific context.

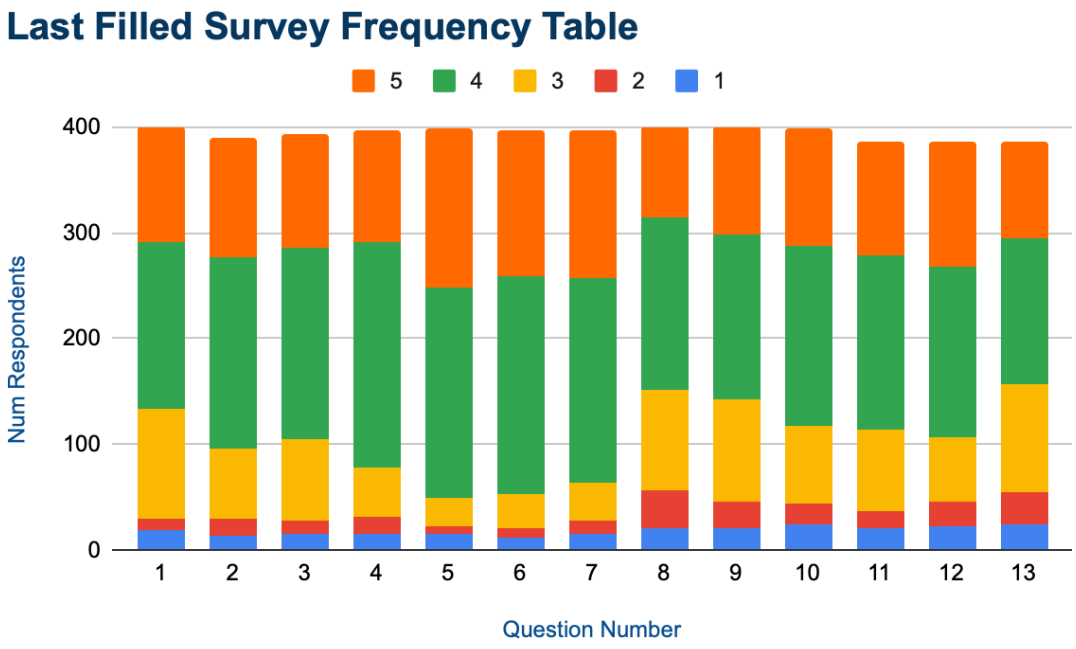
Frequency Tables

How many people gave certain answers to the questions?

Participant's Last Survey Frequency Table

Found the last completed survey by each unique participant ID (post, mid, then pre) aka their final Upower evaluation, and used that to calculate the frequency of answers for each question.

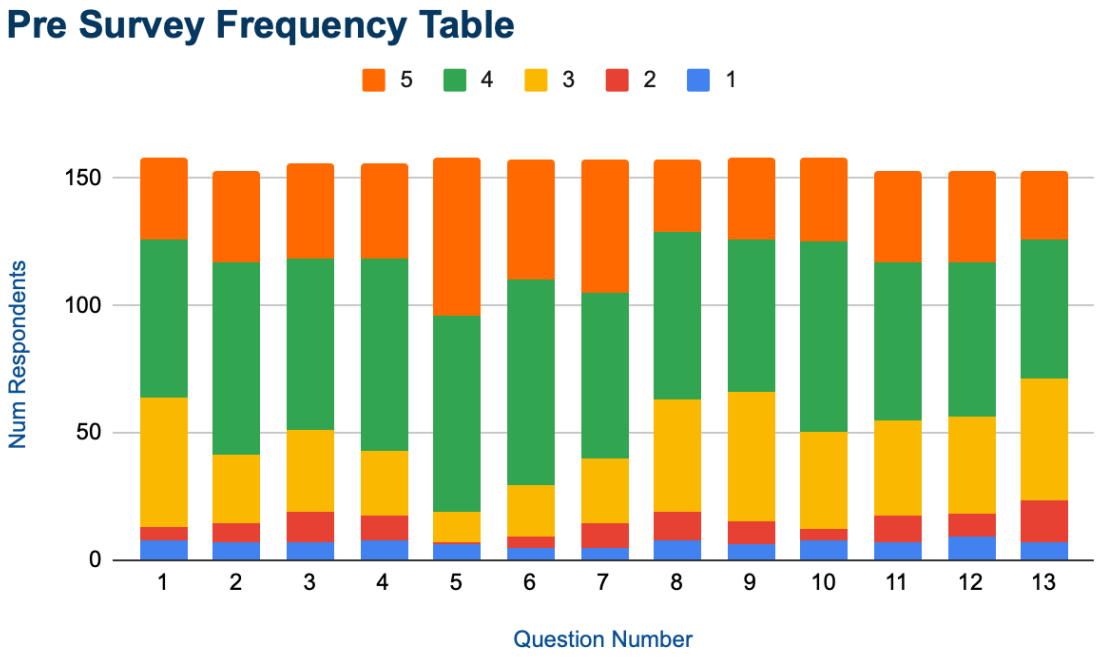
Last Filled Survey Answer Frequency						
question_id	1	2	3	4	5	All
1	19	10	105	158	108	400
2	14	15	66	181	113	389
3	16	12	77	181	107	393
4	16	15	47	214	105	397
5	16	6	27	200	149	398
6	12	9	32	206	138	397
7	16	12	35	195	139	397
8	20	37	95	162	86	400
9	20	25	97	156	102	400
10	25	19	74	170	110	398
11	20	16	77	165	108	386
12	22	24	61	161	119	387
13	25	30	101	139	91	386
All	241	230	894	2288	1475	5128



Most respondents gave 4 or 5s. Differing numbers of total respondents result from some respondents not answering certain questions.

Pre Survey Frequency Table

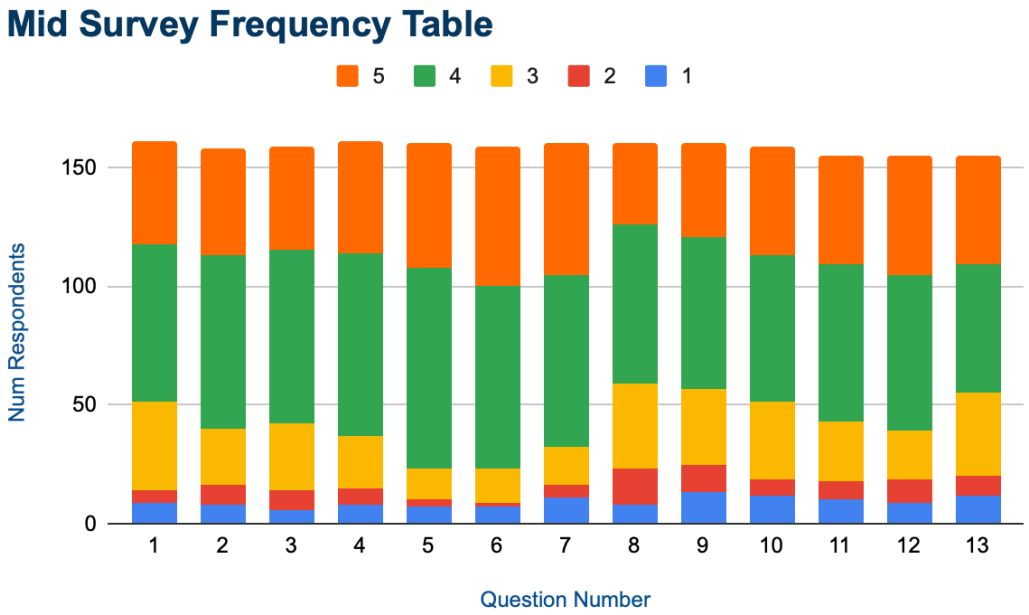
Pre-Survey Answer Frequency						
question_id	1	2	3	4	5	All
1	8	5	51	62	32	158
2	7	7	27	76	36	153
3	7	12	32	67	38	156
4	8	9	26	75	38	156
5	6	1	12	77	62	158
6	5	4	20	81	47	157
7	5	9	26	65	52	157
8	8	11	44	66	28	157
9	6	9	51	60	32	158
10	8	4	38	75	33	158
11	7	10	38	62	36	153
12	9	9	38	61	36	153
13	7	16	48	55	27	153
All	91	106	451	882	497	2027



Mid Survey Frequency Table

Mid-Survey Answer Frequency						
question_id	1	2	3	4	5	All
1	9	5	37	67	43	161
2	8	8	24	73	45	158
3	6	8	28	73	44	159
4	8	7	22	77	47	161
5	7	3	13	85	52	160
6	7	2	14	77	59	159
7	11	5	16	73	55	160
8	8	15	36	67	34	160
9	13	12	32	64	39	160
10	12	7	32	62	46	159
11	10	8	25	66	46	155
12	9	10	20	66	50	155
13	12	8	35	54	46	155

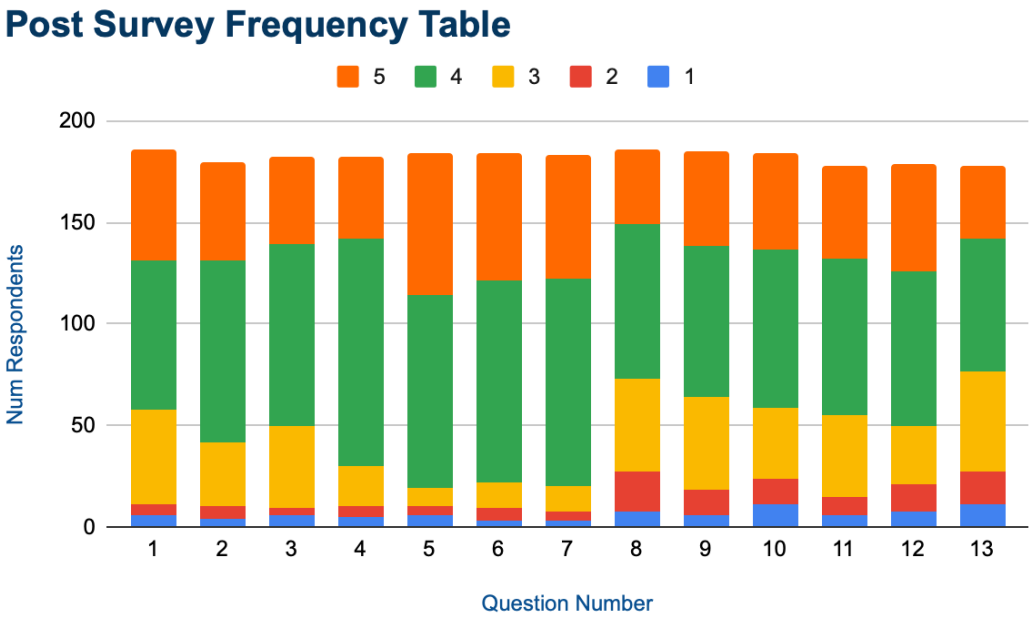
All	120	98	334	904	606	2062



Post Survey Frequency Table

Post-Survey Answer Frequency						
question_id	1	2	3	4	5	All
1	6	5	47	73	55	186
2	4	6	32	89	49	180
3	6	3	41	89	43	182
4	5	5	20	112	40	182
5	6	4	9	95	70	184
6	3	6	13	99	63	184
7	3	5	12	102	61	183
8	8	19	46	76	37	186
9	6	12	46	74	47	185
10	11	13	35	78	47	184
11	6	9	40	77	46	178
12	8	13	29	76	53	179
13	11	16	50	65	36	178

All	83	116	420	1105	647	2371

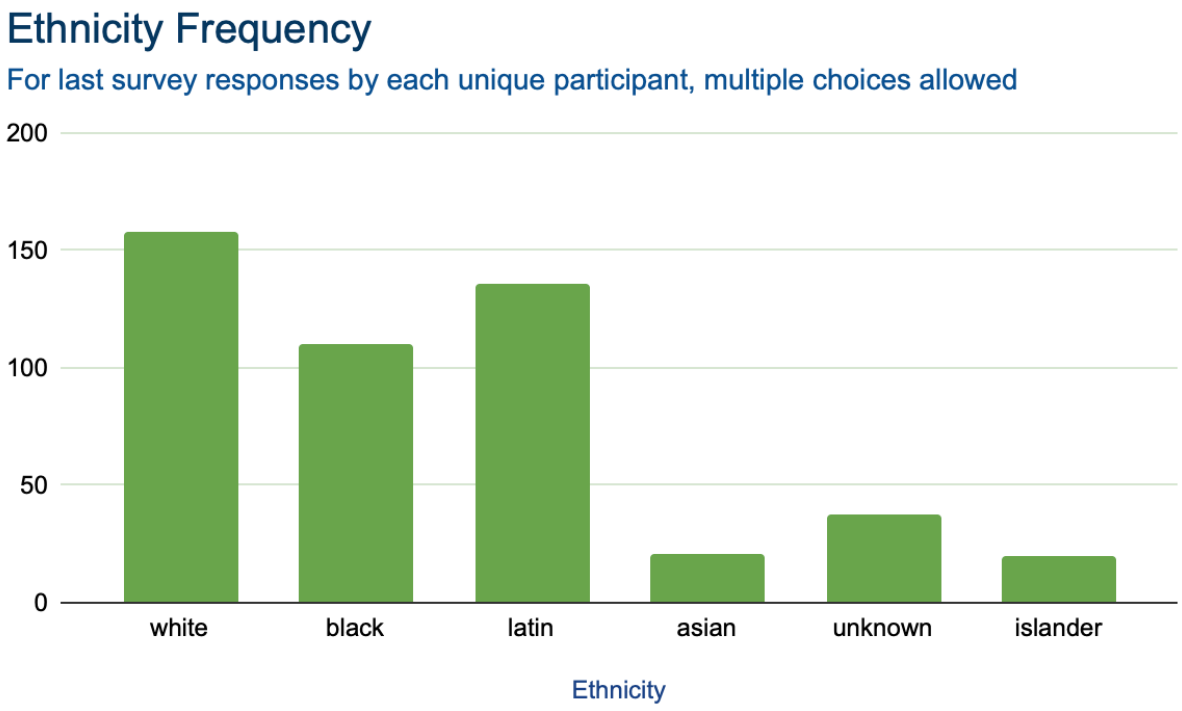


Demographics

Ethnicity

Ethnicity	Count
white	158
black	110
latin	136
asian	21
unknown	38
islander	20

Total responses (participants could choose multiple ethnicities): 483



Respondents were able to choose multiple ethnicities, each label referring to an option on the survey (so refer to the survey to see specific ethnicities that fall under each label).

Used the last response from each unique participant (pre, mid, or post). When inputting data, the first response demographic answers are copy/pasted over to any subsequent spreadsheets.

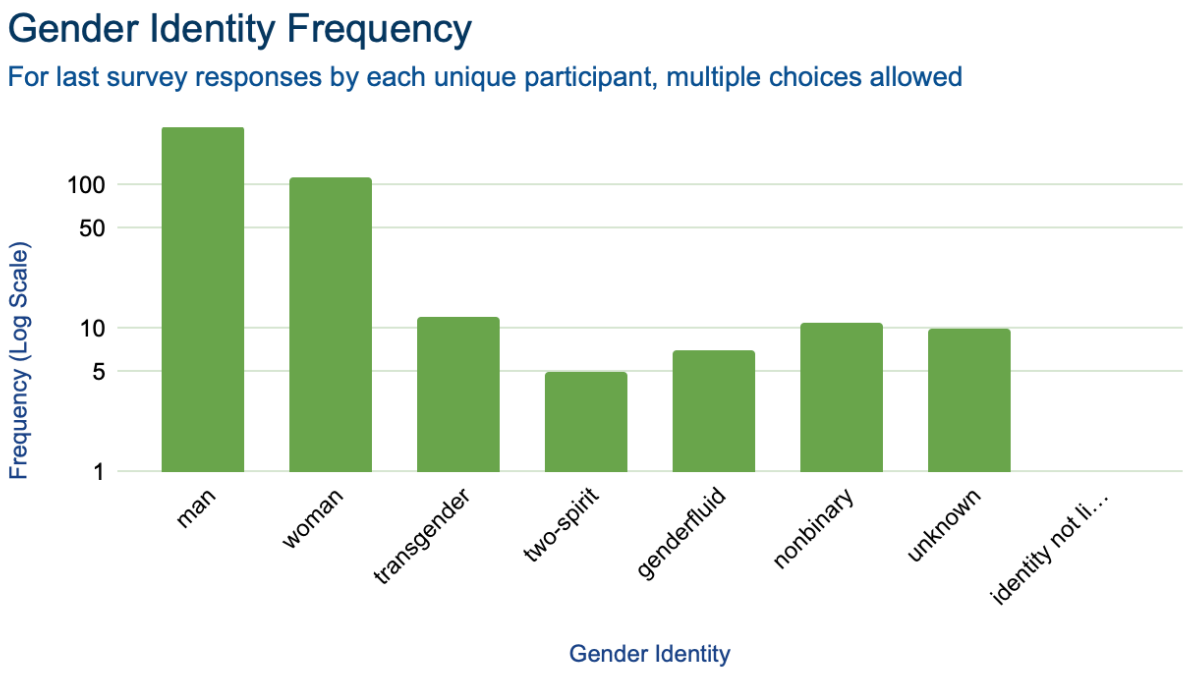
Thus, this graph represents all responses from participants last Upower survey. We can **identify how many participants identified as each ethnic group.** To see **the proportion of respondents that identify as a certain group, compare against the 405 unique participants** (different from the total answers to this question).

Thus a pie chart is not entirely accurate visualization in this case, as the proportions will not add up to 100.

Gender Identity

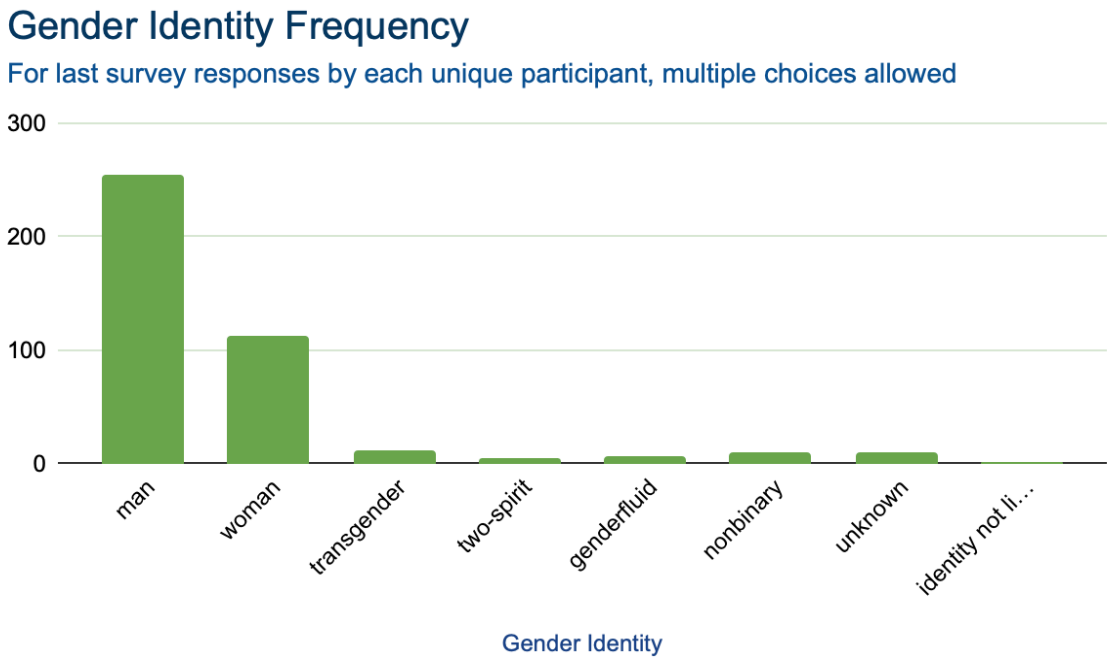
Gender Identity	Frequency	% of total frequency	% of total participants (405)
man	254	0.6150121065	0.6271604938
woman	113	0.2736077482	0.2790123457
transgender	12	0.02905569007	0.02962962963
two-spirit	5	0.01210653753	0.01234567901
genderfluid	7	0.01694915254	0.01728395062
nonbinary	11	0.02663438257	0.02716049383
unknown	10	0.02421307506	0.02469135802
identity not listed	1	0.002421307506	0.002469135802

Total responses (participants could choose multiple): 413



Note the log scale (Y Axis). Man amounts would dwarf all others so this is more descriptive.

Below is a regular scale.



This graph represents all responses from participants last Upower survey. We can **identify how many participants identified as each group.** To see **the proportion of respondents that identify as a certain group, compare against the 405 unique participants** (different from the total answers to this question).

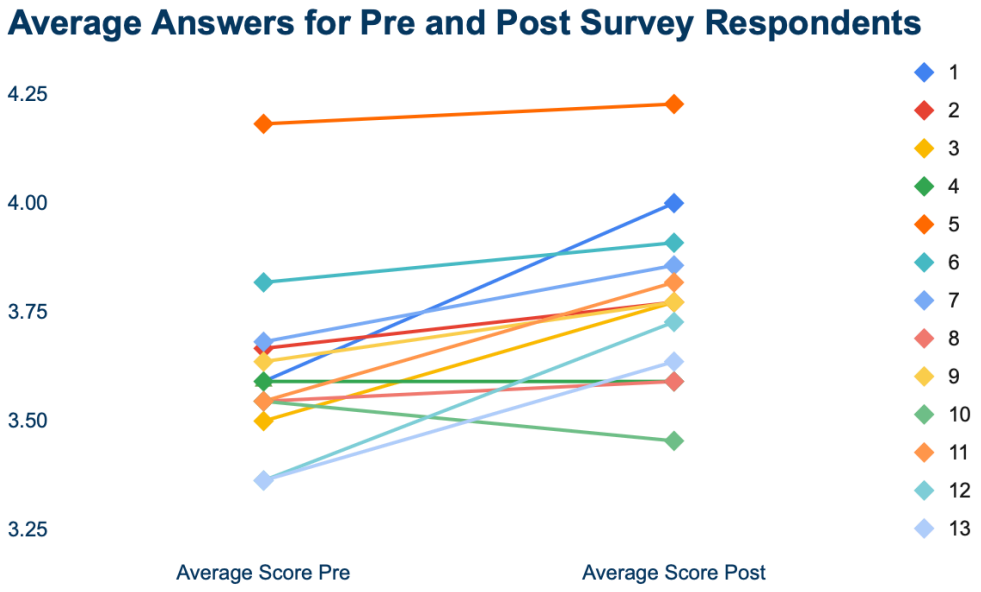
We can see most Upower participants identify as male.

Change In Average Question Score Per Question

Of those who took multiple surveys (and thus we have a before and after), how many have increased scores to questions? (Same answers not counted)

Of those who took Pre and Post

Question Number	Average Score Pre	Average Score Post	Change in Score
1	3.590909091	4	0.4090909091
2	3.666666667	3.772727273	0.1060606061
3	3.5	3.772727273	0.2727272727
4	3.590909091	3.590909091	0
5	4.181818182	4.227272727	0.04545454545
6	3.818181818	3.909090909	0.09090909091
7	3.681818182	3.857142857	0.1753246753
8	3.545454545	3.590909091	0.04545454545
9	3.636363636	3.772727273	0.1363636364
10	3.545454545	3.454545455	-0.09090909091
11	3.545454545	3.818181818	0.2727272727
12	3.363636364	3.727272727	0.3636363636
13	3.363636364	3.636363636	0.2727272727

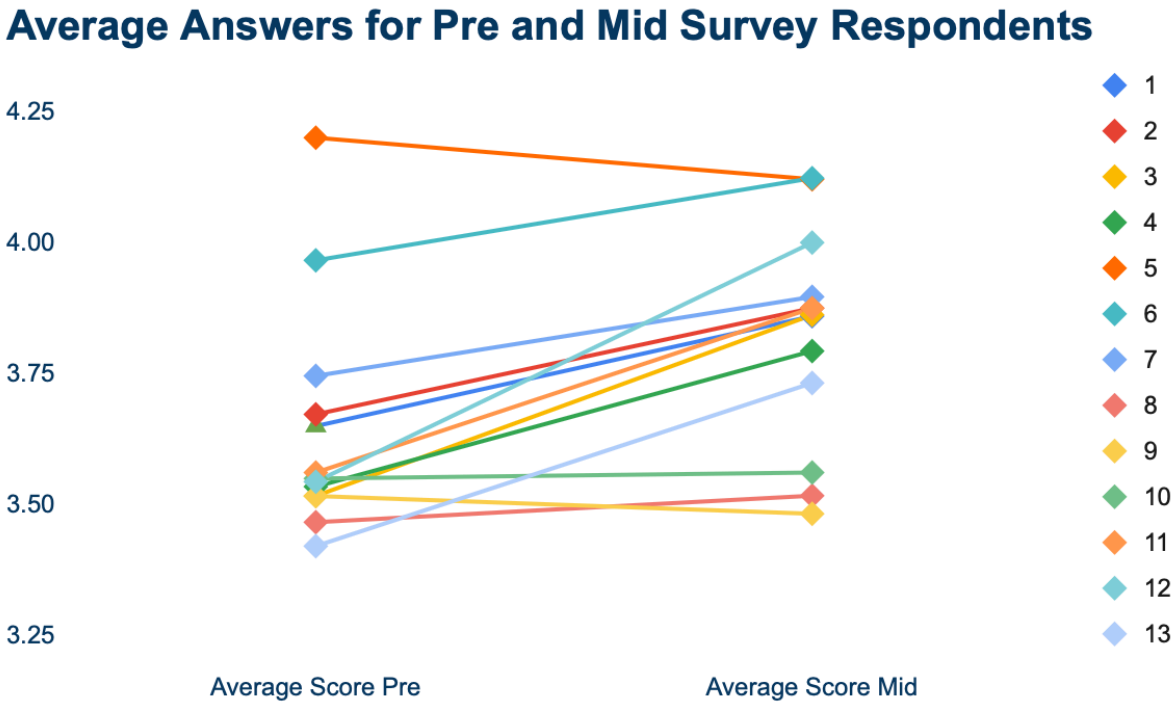


Out of 22 respondents who took both pre and post surveys.

We can see overall growth. Note the small sample size and relatively little response options (just 1, 2, 3, 4, 5). Still, overall responses increase over time in the Upower program.

Of those who took Pre and Mid

Question Number	Average Score Pre	Average Score Mid
1	3.65	3.859649123
2	3.672413793	3.875
3	3.516666667	3.862068966
4	3.534482759	3.793103448
5	4.2	4.120689655
6	3.966101695	4.122807018
7	3.745762712	3.896551724
8	3.466666667	3.517241379
9	3.516666667	3.482758621
10	3.55	3.561403509
11	3.561403509	3.875
12	3.543859649	4
13	3.421052632	3.732142857

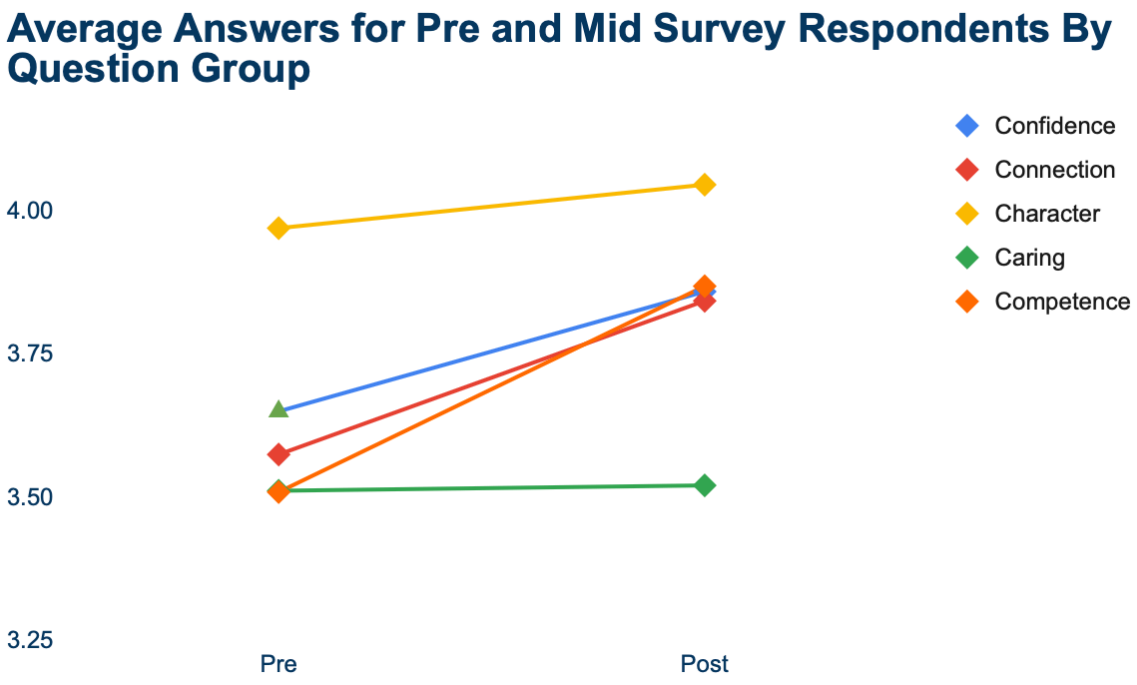


Out of 60 respondents who took both pre and mid surveys.

We can see overall growth. Note the small sample size and relatively little response options (just 1, 2, 3, 4, 5). Still, overall responses increase over time in the Upower program.

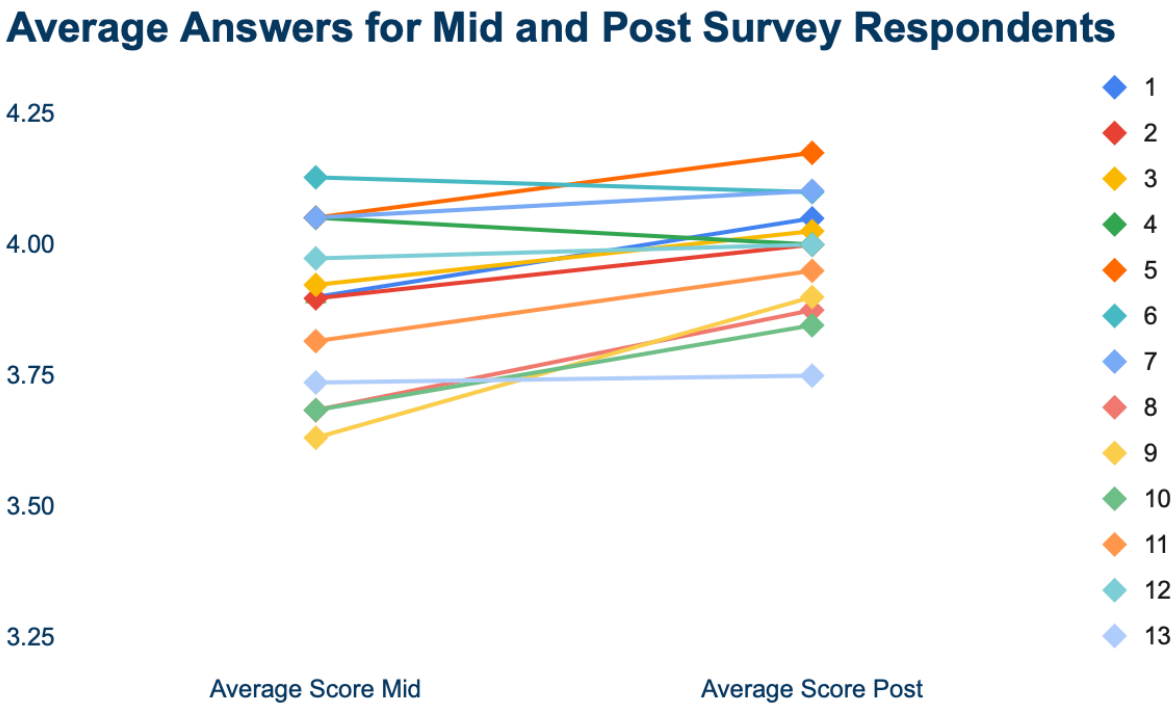
If we look to see how each question category performs:

Question Category	Pre	Post
Confidence	3.65	3.859649123
Connection	3.574521073	3.843390805
Character	3.970621469	4.046682799
Caring	3.511111111	3.520467836
Competence	3.50877193	3.869047619



Of those who took Mid and Post

Question Number	Average Score Mid	Average Score Post
1	3.9	4.05
2	3.897435897	4
3	3.923076923	4.025641026
4	4.051282051	4
5	4.051282051	4.175
6	4.128205128	4.1
7	4.051282051	4.102564103
8	3.684210526	3.875
9	3.631578947	3.9
10	3.684210526	3.846153846
11	3.815789474	3.95
12	3.973684211	4
13	3.736842105	3.75

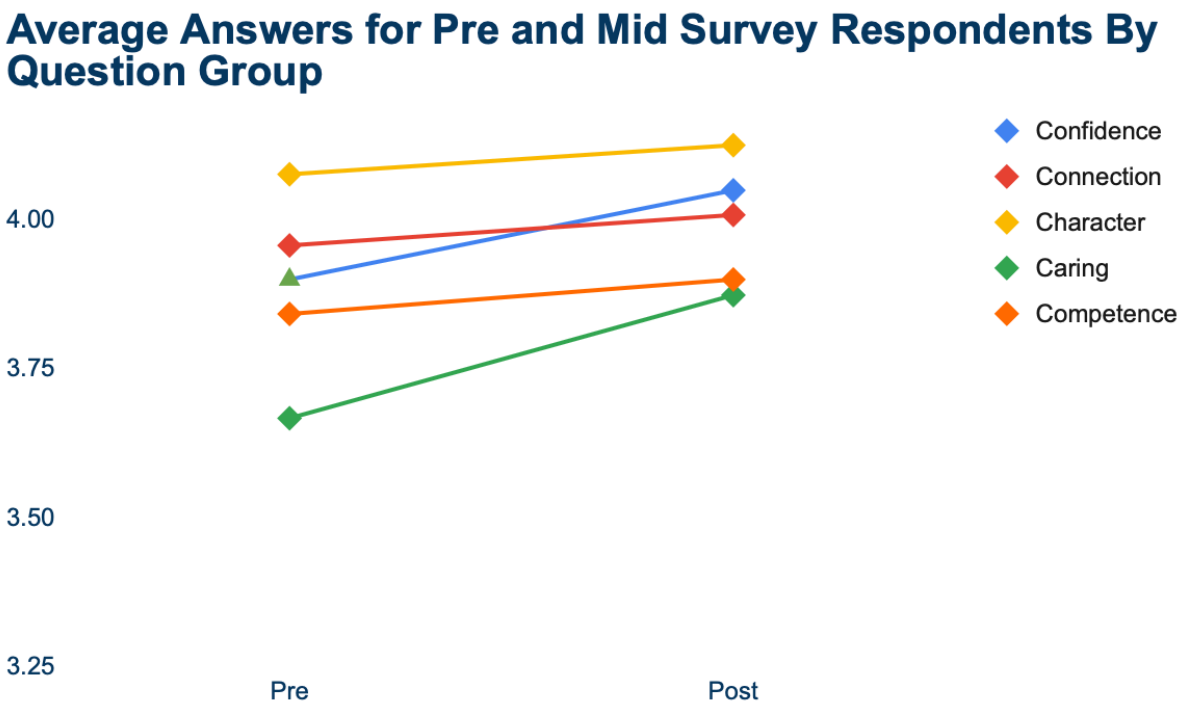


Out of 40 respondents who took both mid and post surveys.

We can see overall growth. Note the small sample size and relatively little response options (just 1, 2, 3, 4, 5). Still, overall responses increase over time in the Upower program.

If we look to see how each question category performs:

Question Category	Pre	Post
Confidence	3.9	4.05
Connection	3.957264957	4.008547009
Character	4.076923077	4.125854701
Caring	3.666666667	3.873717949
Competence	3.842105263	3.9



% Respondents Who Increased Answers In Their Later Surveys

Of all of the people that took 2 surveys (pre & post, pre & mid, mid & post), how many increased answers for the question in their later survey?

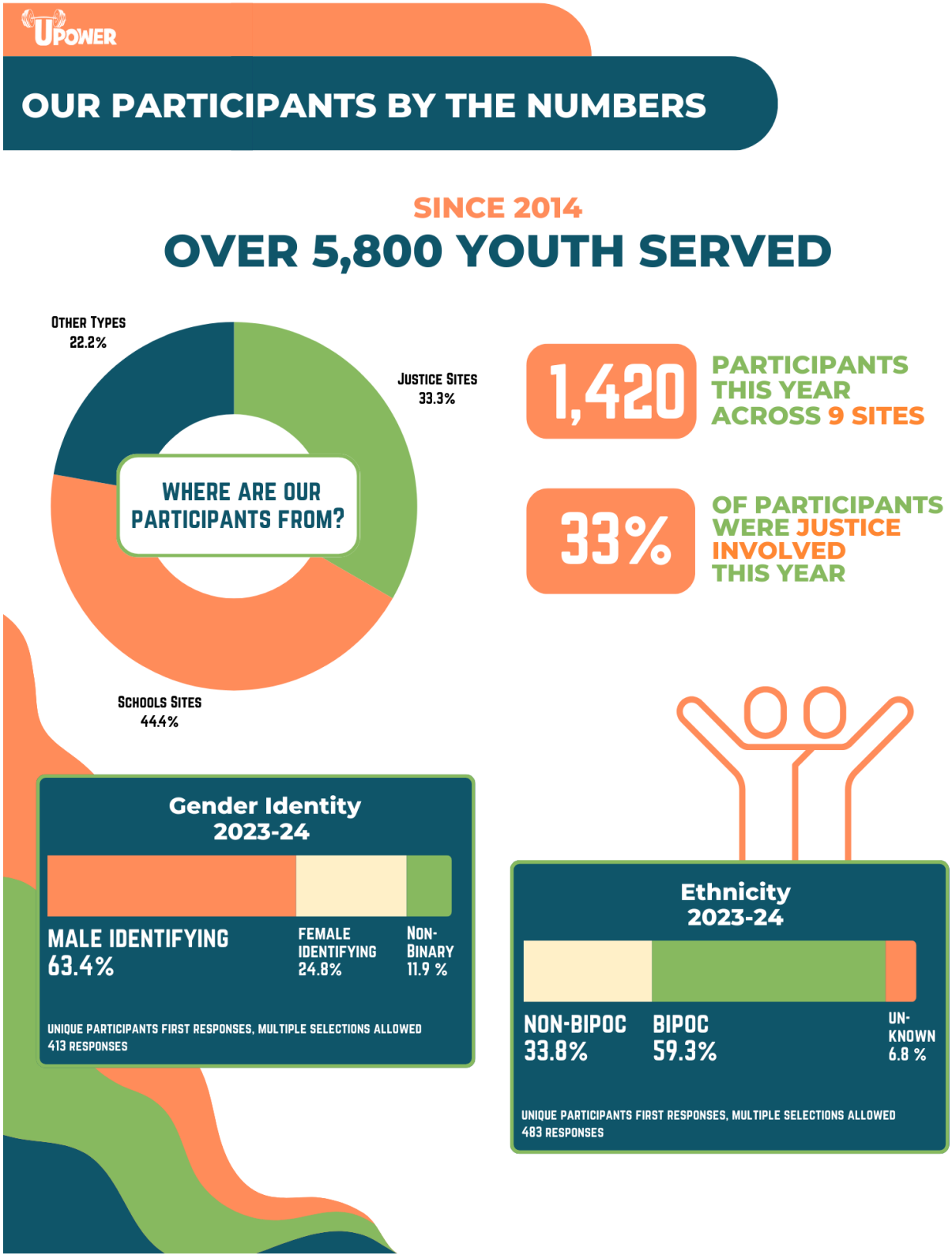
How many respondents took the survey more than once and did they increase their scores per question?					
Question Number	Increased Count pre/post	Increased Count pre/mid	Increased Count mid/post	Total increased for each question	**Percent Increased (total)**
1	9	18	15	42	34.4
2	8	18	9	35	28.7
3	7	26	12	45	36.9
4	6	23	9	38	31.1
5	5	10	12	27	22.1
6	4	16	9	29	23.8
7	7	21	11	39	32
8	6	16	14	36	29.5
9	6	20	15	41	33.6
10	4	16	13	33	27
11	8	26	12	46	37.7
12	9	27	12	48	39.3
13	8	23	13	44	36.1
# Took both survey	22	60	40		
			# multiple surveys taken	122	
			* if a participant took all three, they will be counted for all categories		

To answer questions about Upower's impact on participants, we can compare their survey answers. Did scores increase? How many have increased?

Percentages of interest: comparing before and after Upower programming

Title	Description	Percent	Out of
felt more confident after	Q1	34.4	122
felt better at building positive relations	Q7, 10, 4. Average of their percent increases	30	122
learned to be more inclusive	Q7, 10, 8. Average of their percent increases	31.7	122
strengthened social emotional skills like empathy	Q8, 9, 10. Average of their percent increases	30	122

Designing the Annual Report Spread



MEETING OUR PARTICIPANT'S NEEDS

+ 18,000 ppe masks

900+

Play-empowering items distributed

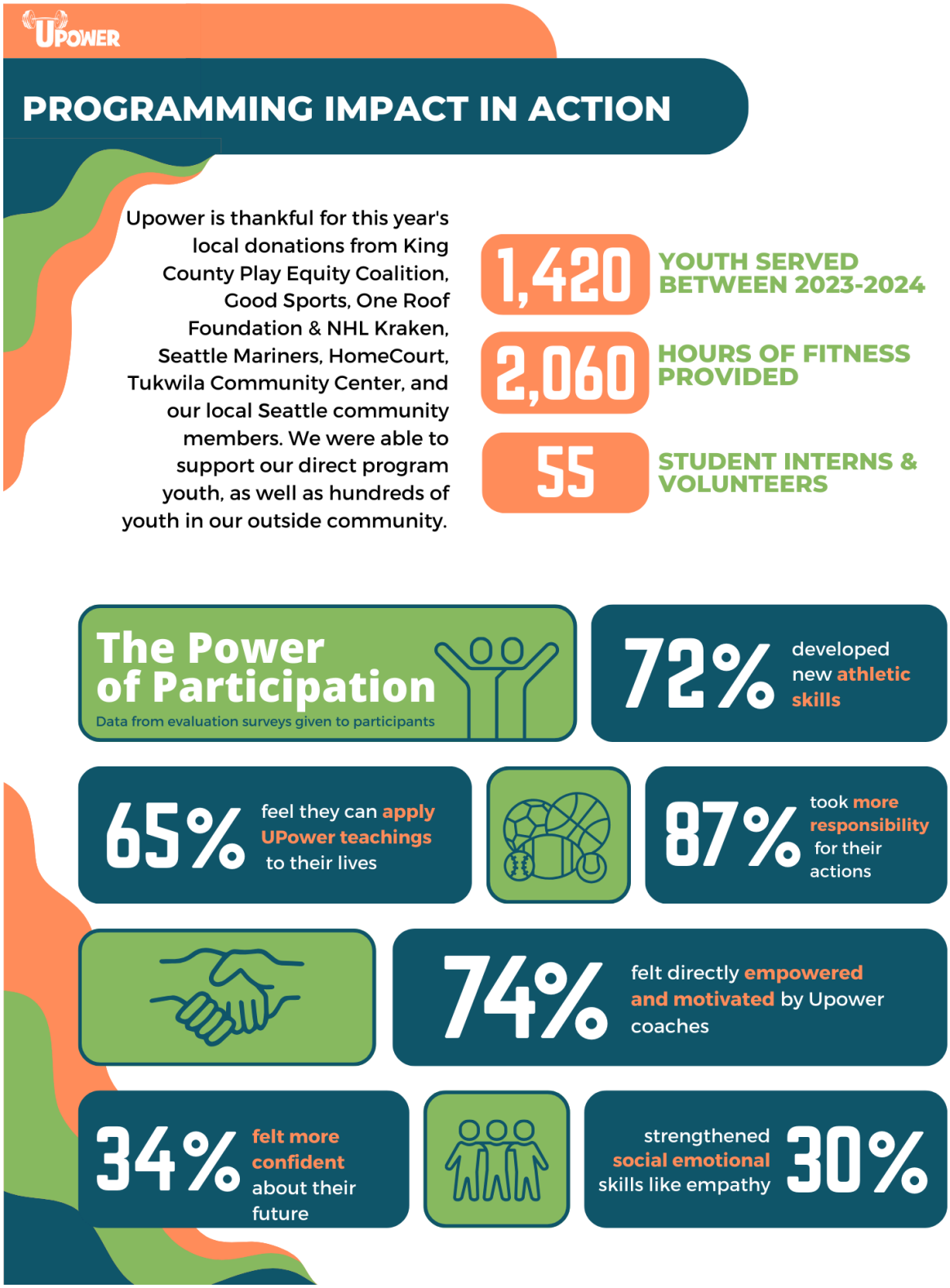
in 2023-2024

Upower is thankful for this year's local donations from King County Play Equity Coalition, Good Sports, One Roof Foundation & NHL Kraken, Seattle Mariners, HomeCourt, Tukwila Community Center, and our local Seattle community members. We were able to support our direct program youth, as well as hundreds of youth in our outside community.

Donate Today

& give kids the ability to play!

www.u-power.org/donate-now/



*note the text copy and qr code are currently placeholders

Learnings & Skills Developed

In the technical realm, I strengthened my skills in excel (sheets equivalent) and google colabratory python (pandas library). Working with real data is an invaluable experience as a student as it is realistically messy and open-ended. Additionally, this project had a unique freedom in that there were no defined set of steps to take. Last year's survey was so different, and the analysis not particularly applicable that it was up to me and my fellow interns to design a system for analyzing this data. That freedom taught me a lot about descriptive statistics and how to start from scratch and not just plug in numbers into a predefined analysis set-up.

In the non-technical realm, adaptability, communication, time management, and problem solving are the skills I developed most. The freedom to conduct the analysis however I saw fit gave me the chance to be collaborative by engaging in discussions with my fellow interns, data and math professors at Seattle University, and Upower staff. As the head intern of this project, coordinating communication between us interns and splitting the workload was important to productivity. Problem solving came in the form of deciding on avenues of analyzing the data and implementing those solutions. Any assumptions made were made with thought, reasoning, and likely after discussion.

Next Steps & Recommendations

Survey Design

I. Keeping the survey design comparable would allow for Upower to compare next year's analysis to this years, etc.
II. Based on the percentages we calculated in this project, ensure these are the types of questions we want to answer. If we want to evaluate the importance of physical activity to mental health, for example, we should ensure a question or two like that exist on the survey.
III. Keep it short

Analysis

I. You could also analyze by site, also by demographics. ex. How do sites differ in their answers to questions? Does ethnicity impact respondents' view on Upower?
II. Some other questions I brainstormed at the beginning of the project:
    A. Distribution of responses across genders, ethnicities, languages? (Is Upower more impactful to different groups?) (response bias- can certain people better understand
    B. Chi-square test: Examine the association between categorical variables (e.g., question response and participant demographics).
    C. Spearman's rank correlation coefficient: Determine the strength and direction of the monotonic relationship between two ranked variables.d questions or more willing to fill out the survey?)

Data Input/Processing

The conflicts arise from wanting a method of input that makes sense and isn't overwhelming for humans, and a spreadsheet that is optimal for data analysis. These do not often mix. See wide vs long tables (what I converted the raw sheet to).

For example, having separate sub sheets for each site and for each time survey is more digestible for people but not helpful when we want to compare these things with data analysis softwares. Thus our need to clean the data pretty heavily. In the future:

I. To cut down on need to clean data after and change it before analyzing too much, could:
    A. Make a google form to submit each, and have google convert it to a nice sheet format. Alternatively, another survey software could do this? Requires set up and manual input but more uniform as demographic categories are enforced.
    B. Input into sheets manually, but enforce standard labels for demographics, etc. could also use drop down menus and other data validation tools
II. Have one big spreadsheet for pre, one for mid, one for post. Have a column specify the site. Could link it to update a mass spreadsheet if we want all the data.

Automating the Process

I. Could use my colab script, develop it further so it is easier for a non programmer to run against a raw data file (standardized inputs, etc). Then export that .csv file, import to sheets and have pivot tables and auto generated charts.

Annual Report

I. Edit/update the blurbs of text. Guidance on what topics would be covered provided.
II. Update the QR Code. Maybe make a custom shaped one!
III. I made 3 pages with the thought that the 4th would be the Train the Trainer data driven outcomes page.
IV. The change over time percentages are lower, could change to just use the last taken survey data but if the text implies it is over time improvement, the data calculation should reflect that

Conclusion

The goals of this project were met, deliverables were completed and I value the experiential learning I gained. This project allowed me to not only apply my interests in data science, but also my problem solving, professional communication, and design skills.

A collaborative effort between me, my Upower mentors, and my fellow data interns was crucial to my work and learning. A deep understanding of the Upower experience and the context the data comes from is a crucial influence on how one analyzes the data- an accurate story cannot be told in an inaccurate context.

For example, knowing that paper copies of these surveys are distributed to an entirely youth audience, from a diverse socio-economic population in potentially traumatic life situations is important to know. We must work with handwritten names, making it difficult to identify recurring participants at times. Not everyone is entirely willing or capable of accurately reflecting on their experience in a survey. A further example, past improvements of the survey were to shorten it and to remove questions that demotivated or confused respondents. As always, statistics and graphs cannot tell the entire story- the context matters. I am grateful to my Upower mentors and coaches for answering any questions I had and educating me in trauma-informed coaching and the Upower experience- crucial contexts to successfully working with this data.

In addition, learning from other data interns from business analytics and pure data science backgrounds was very informative. Transitioning from more individual classwork to a collaborative effort is an invaluable skill I will apply to future work in industry.

Source Code and Spreadsheets

See attached files or links to my github (@cmdenault).

References & Such

Professor Sloughter and Professor Mendible (Seattle University) for help ideating approaches
NonProfit Data Visualizations PDF
Likert Scale Best Practices
Non Profit Visuals Inspiration