

Final Project

AUTHOR
Sophie Ryan

PUBLISHED
August 16, 2024

National Wellbeing Survey, United States, 2021

The National Wellbeing Survey (NWS) is a population-based survey of adults ages 18-64 in the United States. The purpose of this study was to develop an understanding of overall wellbeing. Participants were recruited online, and the sample is comprised of 4,041 adults in the United States. The dataset includes individual demographic measures such as age, race, and gender. The variables regard physical and mental health, social relationships, community involvement, socioeconomic status, COVID-19 experiences, employment experiences, and political orientation. For my project, I am mainly interested in investigating how employment satisfaction relates to mental health, life satisfaction, and well-being in general. The original dataset consisted of 4,041 observations (one for each participant in the study), and 254 variables. I narrowed down this data to 37 variables that I was particularly interested in analyzing and were most pertinent to my research questions.

<https://www.icpsr.umich.edu/web/ICPSR/studies/38879/summary>

```
library(tidyverse)
```

Warning: package 'ggplot2' was built under R version 4.3.2

```
— Attaching core tidyverse packages —————
tidyverse 2.0.0 —
✓ dplyr      1.1.3      ✓ readr      2.1.4
✓ forcats    1.0.0      ✓ stringr    1.5.0
✓ ggplot2    3.5.1      ✓ tibble     3.2.1
✓ lubridate  1.9.2      ✓ tidyr      1.3.0
✓ purrr      1.0.2
— Conflicts —————
tidyverse_conflicts() —
✖ dplyr::filter() masks stats::filter()
✖ dplyr::lag()     masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>)
to force all conflicts to become errors
```

```
library(readxl)
library(haven)
library(stringr)
library(ggplot2)

#Read the original data and viewed the dimensions.
wellbeing_data <- read_dta("38879-0001-Data.dta")
dim(wellbeing_data)
```

```
[1] 4014 254
```

```
#Looked at the first few rows of the original dataframe
head(wellbeing_data)
```

```
# A tibble: 6 × 254
  SURVEYID ENDDATE   AGE   SEX   HISPANIC R_WHITE
R_BLACK R_AIAN  R_API
  <chr>    <chr>    <dbl> <dbl> <dbl> <dbl>
<dbl> <dbl> <dbl>
1 2021_933 04-FEB-2021 36     2 [Fem... 1 [Yes]  0 [Ite... 0
[Ite... 0 [Ite... 0 [Ite...
2 2021_1336 06-FEB-2021 47     2 [Fem... 2 [No]   0 [Ite... 0
[Ite... 0 [Ite... 1 [Ite...
3 2021_733 03-FEB-2021 51     2 [Fem... 2 [No]   0 [Ite... 1
[Ite... 0 [Ite... 0 [Ite...
4 2021_2880 22-FEB-2021 45     2 [Fem... 2 [No]   1 [Ite... 0
[Ite... 0 [Ite... 0 [Ite...
5 2021_3971 16-MAR-2021 36     1 [Mal... 2 [No]   1 [Ite... 0
[Ite... 0 [Ite... 0 [Ite...
6 2021_3279 08-MAR-2021 52     1 [Mal... 2 [No]   0 [Ite... 0
[Ite... 0 [Ite... 1 [Ite...
# i 245 more variables: R_OTHER <dbl>, R_DK <dbl>,
RACEREC <dbl>,
# STATE_NAME <chr>, STFIPS <chr>, COUNTY_NAME <chr>, FIPS
<chr>,
# COUNTY_STATE <chr>, RUCC <dbl>, HAPPY <dbl>,
LAD_NOW <dbl>,
# LAD_FUT <dbl>, LAD_REL <dbl>, CHILD_BETTER
<dbl>,
# SAT1 <dbl>, SAT2 <dbl>, SAT3 <dbl>, SAT4
<dbl>,
# SAT5 <dbl>, HOPEFUL <dbl>, MEANING <dbl>, RES1
<dbl>,
# RES2 <dbl>, RES3 <dbl>, RES4 <dbl>, RES5
<dbl>, ...
```

Here we can see the data before it was transformed, consisting of 4041

rows and 254 columns. I cleaned and tidied the data, narrowing it down to 37 columns, and recoded variables to be mainly categorical rather than numeric or binary. I created a new csv file with the transformed data.

```
#Reading the cleaned data.
new_wellbeing <- read_csv("new_wellbeing.csv")
```

Rows: 4014 Columns: 37
— Column specification

```
Delimiter: ","
chr (33): SURVEYID, SEX, Race, HISPANIC, HAPPY, Life
Satisfaction, LONELY1, ...
dbl (4): AGE, Life Position - Scale 1:10, Future Position -
Scale 1:10, SVY...
```

```
i Use `spec()` to retrieve the full column specification for
this data.
i Specify the column types or set `show_col_types = FALSE` to
quiet this message.
```

```
#Dropping NA values for the columns I am interested in
new_wellbeing <- new_wellbeing |>
  drop_na(WORK_HRS, MENT_HEALTH, `Fulfilling Work`, HAI)

#Finding dimensions.
dim(new_wellbeing)
```

```
[1] 2417 37
```

```
#Look at the first few rows of the transformed data.
head(new_wellbeing)
```

```
# A tibble: 6 × 37
  SURVEYID  AGE SEX Race HISPANIC HAPPY `Life
Satisfaction`
  <chr>    <dbl> <chr> <chr> <chr>    <chr>    <chr>
1 2021_933 36 female Other yes    very_happy satisfied
2 2021_1336 47 female Asian no    rather_happy satisfied
3 2021_733 51 female Black no    rather_happy satisfied
4 2021_2880 45 female White no    rather_happy satisfied
5 2021_3971 36 male White no    very_happy satisfied
6 2021_3279 52 male Asian no    rather_happy satisfied
# i 30 more variables: `Life Position - Scale 1:10` <dbl>,
# `Future Position - Scale 1:10` <dbl>, LONELY1 <chr>,
RELIG_IMP <chr>,
```

```
# SOCIAL <chr>, PHYS_HEALTH <chr>, OVERWEIGHT <chr>, PTSD
<chr>,
# DEPRESSION <chr>, ANXIETY <chr>, MENT_HEALTH <chr>,
ALCOHOL <chr>,
# SMOKER <chr>, MJSMOKE_EVER <chr>, EMP_FULL1 <chr>, LAIDOFF
<chr>,
# STUD_FT <chr>, SAT_WORK <chr>, WORK_HRS <chr>, `Manageable
Workload` <chr>,
# `Fulfilling Work` <chr>, `Covid Impact` <chr>, `Tested
Positive` <chr>, ...
```

Exploratory Data Analysis

Now that the data is more succinct and easy to read, I began to conduct analyses and statistics.

```
#Creating a function that finds range and mean of a given variable
my_function_wellbeing <- function(x) {
  list(range = range(x, na.rm = T), mean = mean(x, na.rm = T))
}

#Summarizing the age column.
print(my_function_wellbeing(new_wellbeing$AGE))
```

```
$range
[1] 18 64
```

```
$mean
[1] 39.13198
```

```
#Summarizing each variable of interest using group_by and summarise
new_wellbeing |>
  group_by(WORK_HRS) |>
  summarize(count = n()) |>
  arrange(desc(WORK_HRS))
```

```
# A tibble: 5 × 2
  WORK_HRS count
  <chr>      <int>
1 refuse         3
2 <20         271
3 45+          323
4 35-45       1196
5 20-34        624
```

```
new_wellbeing |>
  group_by(MENT_HEALTH) |>
  summarize(count = n()) |>
  arrange(desc(MENT_HEALTH))
```

```
# A tibble: 6 × 2
  MENT_HEALTH count
  <chr>         <int>
1 very good     743
2 poor          105
3 n/a           2
4 good          624
5 fair          331
6 excellent     612
```

```
new_wellbeing |>
  group_by(`Fulfilling Work`) |>
  summarize(count = n()) |>
  arrange(desc(`Fulfilling Work`))
```

```
# A tibble: 6 × 2
  `Fulfilling Work` count
  <chr>               <int>
1 strongly disagree  118
2 strongly agree     610
3 somewhat disagree  275
4 somewhat agree     826
5 neutral            585
6 n/a                 3
```

```
new_wellbeing |>
  group_by(HAPPY) |>
  summarize(count = n()) |>
  arrange(desc(HAPPY))
```

```
# A tibble: 4 × 2
  HAPPY          count
  <chr>          <int>
1 very_happy     864
2 rather_happy  1186
3 not_very_happy 318
4 not_at_all     49
```

```
new_wellbeing |>
  group_by(LONELY1) |>
  summarize(count = n()) |>
```

```
arrange(desc(LONELY1))
```

```
# A tibble: 5 × 2
```

	LONELY1	count
	<chr>	<int>
1	unsure	43
2	sometimes lonely	845
3	rarely lonely	646
4	never lonely	567
5	always lonely	316

```
new_wellbeing |>
  group_by(EDUC) |>
  summarize(count = n()) |>
  arrange(desc(EDUC))
```

```
# A tibble: 6 × 2
```

	EDUC	count
	<chr>	<int>
1	some college	431
2	post-bachelors	467
3	did not complete high school	53
4	bachelors degree	702
5	associates degree	272
6	GED	492

```
new_wellbeing |>
  group_by(`Manageable Workload`) |>
  summarize(count = n()) |>
  arrange(desc(`Manageable Workload`))
```

```
# A tibble: 6 × 2
```

	`Manageable Workload`	count
	<chr>	<int>
1	strongly disagree	83
2	strongly agree	754
3	somewhat disagree	204
4	somewhat agree	885
5	neutral	488
6	n/a	3

Individual Variable Summaries

The age of the sample ranges from 18 to 64, with a mean age of 39.13.

The most common range of hours worked per week is 35-45 hours, with a count of 1196 respondents. The second most common amount of working

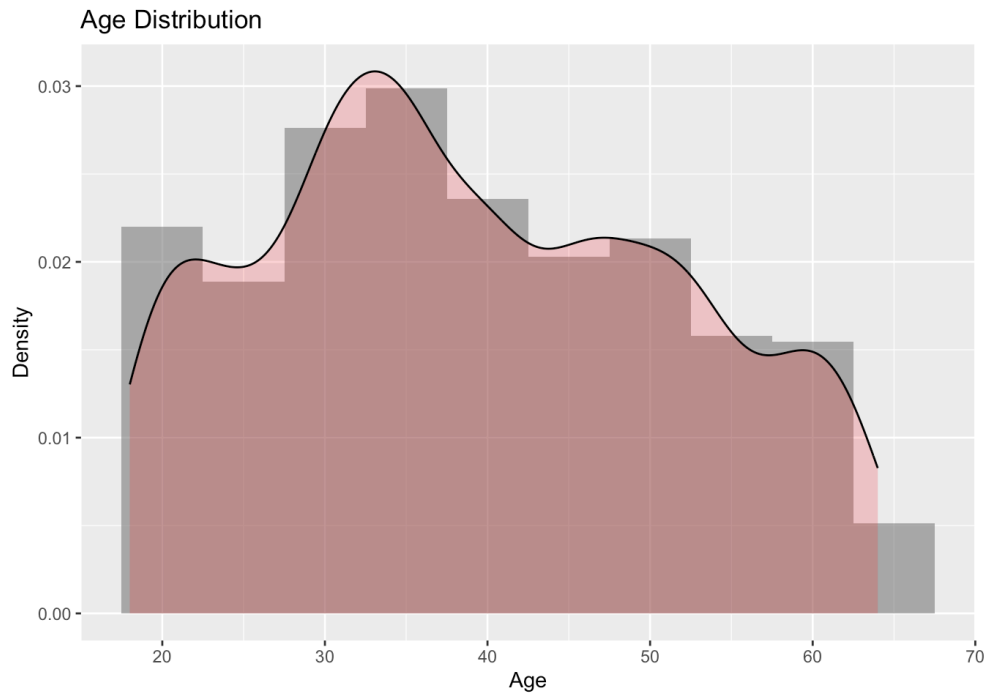
hours is 20-34. The most common answer for state of mental health is "very good", with a count of 743. Most answers consisted of "good", "very good", or "excellent", and the count for "poor" mental health was only 105. For the Fulfilling Work column, the most common answer was "somewhat agree", with a count of 826. "Strongly disagree" was the least common answer with a count of 118. Responses about happiness were most commonly "rather happy", with a count of 1186. "Not at all happy" was the least voted, with a count of 49. The distribution of loneliness answers was a little more spread out, with "sometimes lonely" most common at 845 counts and "always lonely" least common at 316. In the education column, the most common degree obtained is a bachelors degree, with a count of 702. Finally, most respondents voted that they "somewhat agree" that their workload is manageable, with a count of 885. These statistics can help us gauge how the majority of people are responding on individual measures. In general, people seem to be averagely satisfied with life and work. Most responses were somewhere in the middle, and we never see majority votes for either ends of the extreme on any given measure.

Visualizations

First I created a histogram to see the age distribution among participants.

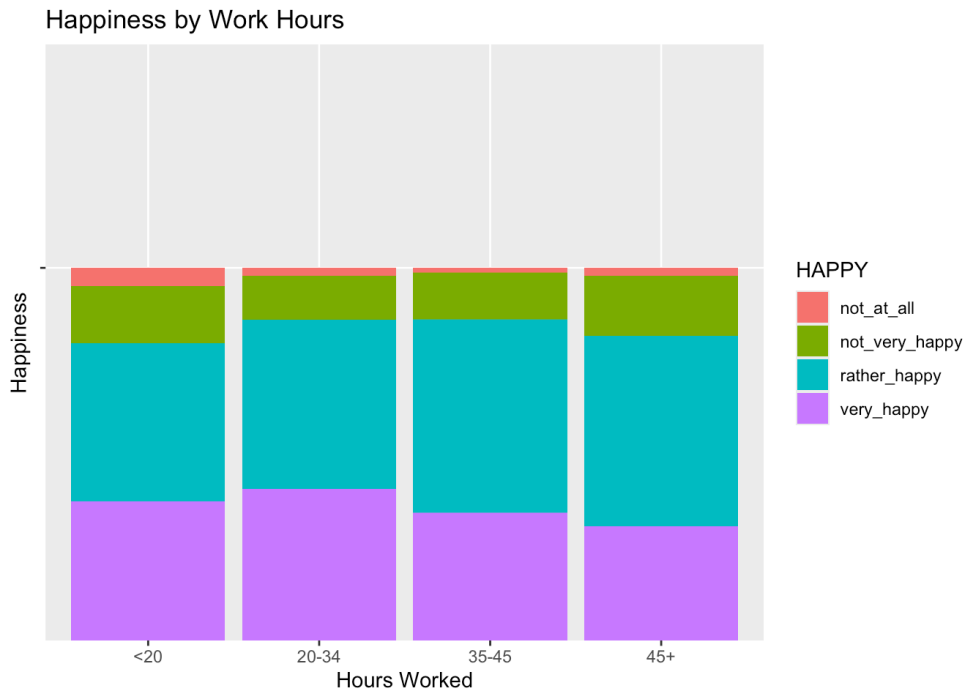
```
ggplot(new_wellbeing, aes(x = AGE)) +  
  geom_histogram(aes(y = ..density..), binwidth = 5, alpha = 0.2, fill = "red") +  
  geom_density(alpha = 0.2, fill = "red") +  
  labs(title = "Age Distribution", x = "Age", y = "Density")
```

Warning: The dot-dot notation (`..density..`) was deprecated in ggplot2 3.4.0.
i Please use `after_stat(density)` instead.



How do the amount of hours worked correlate with reported levels of happiness?

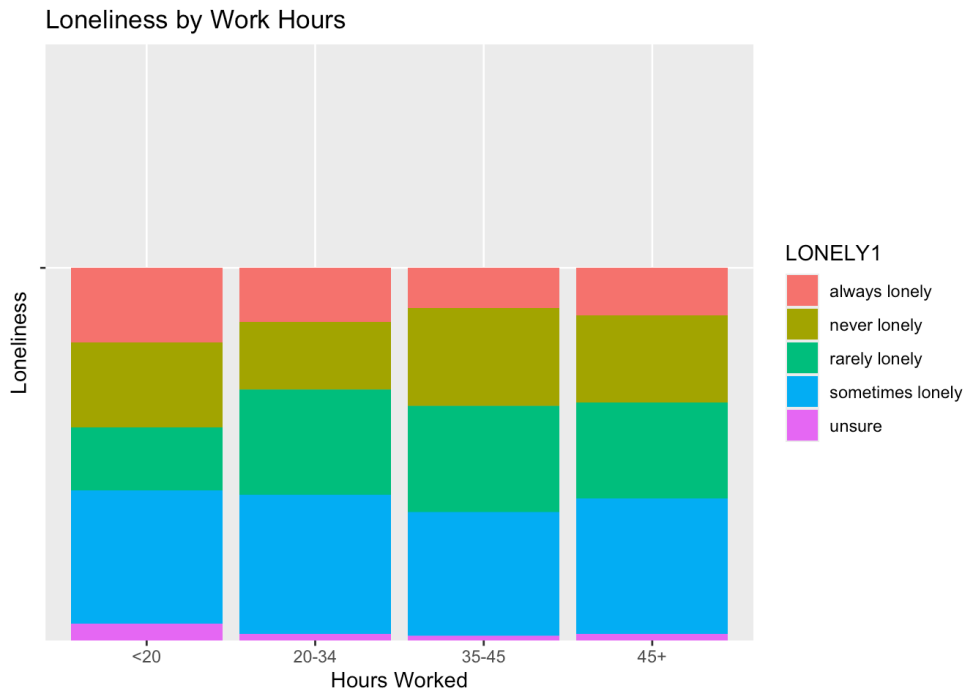
```
new_wellbeing |>
  subset(!is.na(WORK_HRS)) |>
  subset(WORK_HRS != "refuse") |>
  ggplot(aes(fill = HAPPY, x = WORK_HRS, y = "")) +
  geom_bar(stat = "identity", position = "fill") +
  labs(title = "Happiness by Work Hours", x = "Hours W
```

This graph shows us some patterns in how weekly hours at work correlate with happiness. We can see that the range of hours in which the most people reported being "very happy" is 20-34. The least amount of "very happy" responses fall in the 45+ hours range. The graph also shows that the majority of people who reported being "not very happy" or "not at all happy" fall at one of two ends of the spectrum; working either under 20 or over 45 hours per week.

How do the amount of hours worked correlate with reported levels of loneliness?

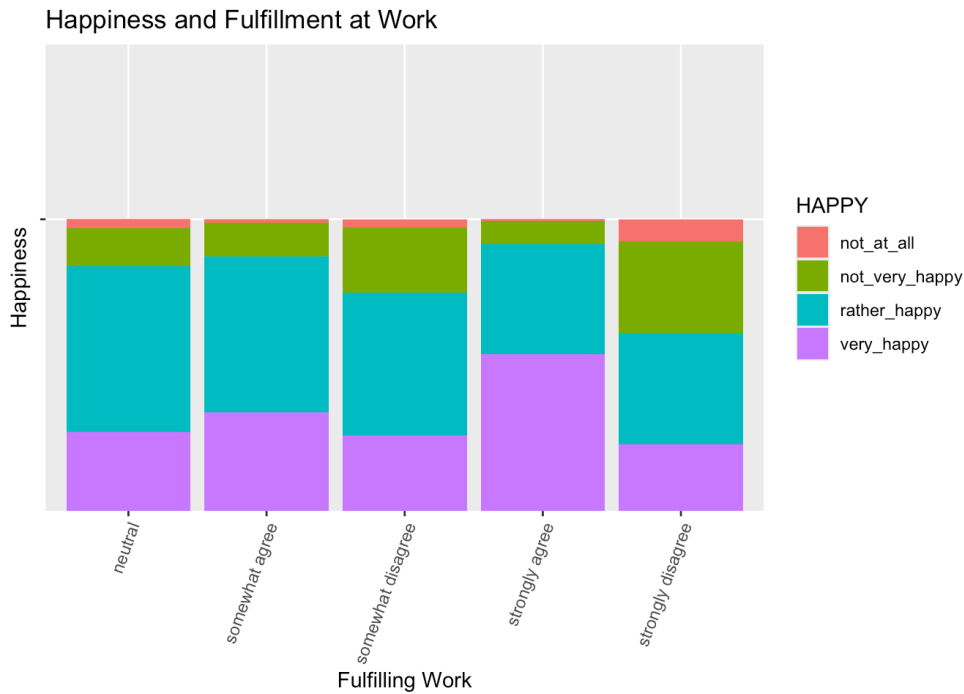
```
new_wellbeing |>
  subset(!is.na(WORK_HRS)) |>
  subset(WORK_HRS != "refuse") |>
  ggplot(aes(fill = LONELY1, x = WORK_HRS ,y = "")) +
  geom_bar(stat = "identity", position = "fill") +
  labs(title = "Loneliness by Work Hours", x = "Hours \
```



This graph shows us some patterns in people's levels of loneliness. The largest proportion of "always lonely" responses fall into the under 20 hours worked category. The majority of "never lonely" responses fall in the 35-45 hours category. It seems that fewer work hours correlates with more loneliness in general.

Does a fulfilling job correlate with more happiness?

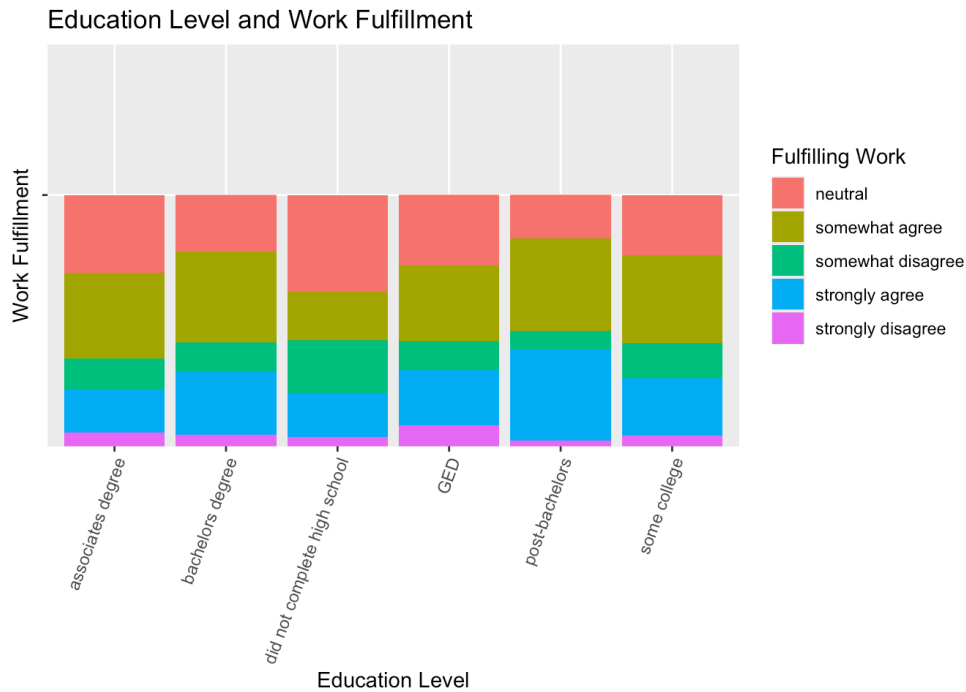
```
new_wellbeing |>
  subset(!is.na(`Fulfilling Work`))|>
  subset(`Fulfilling Work` != "n/a") |>
  ggplot(aes(fill = HAPPY, x = `Fulfilling Work`, y = `Happiness`)) +
  geom_bar(stat = "identity", position = "fill") +
  labs(title = "Happiness and Fulfillment at Work", x = "Fulfilling Work", y = "Happiness") +
  theme(axis.text.x=element_text(angle=70, hjust=1))
```



This graph can give us insight on correlations between a fulfilling job and levels of happiness. The results clearly show that those who reported they "strongly agree" their work is fulfilling also have by far the highest proportion of "very happy" responses. Those who "strongly disagree" that their work is fulfilling have both the lowest frequency of "very happy" and the highest frequencies of "not very happy" and "not at all happy." Additionally, those who reported "somewhat disagree" clearly have the second highest amount of "not at all" and "not very happy" reports. This shows a correlation between work satisfaction and overall satisfaction in life.

How does level of education correlate with a fulfilling job?

```
new_wellbeing |>
  subset(!is.na(`Fulfilling Work`))|>
  subset(`Fulfilling Work` != "n/a") |>
  ggplot(aes(fill = `Fulfilling Work`, x = EDUC ,y = ""
  geom_bar(stat = "identity", position = "fill") +
  labs(title = "Education Level and Work Fulfillment",
  theme(axis.text.x=element_text(angle=70, hjust=1))
```



This graph can show us some correlations between level of education respondents received and how fulfilling their jobs are. Those who have only received a high school diploma/GED had the most reports of "strongly disagree" that their work is fulfilling, while those who have completed post-bachelors have the least amount. Those with a post-bachelors degree also clearly report the highest proportion of "strongly agree". Those who did not complete high school have the highest proportion of "neutral" responses.

How does a manageable workload correlate with mental health?

```
new_wellbeing |>
  subset(!is.na(`Manageable Workload`))|>
  subset(`Manageable Workload` != "n/a") |>
  ggplot(aes(fill = MENT_HEALTH, x = `Manageable Workload`))|>
  geom_bar(stat = "identity", position = "fill") +
  labs(title = "Workload and Mental Health", x = "Manageable Workload") +
  theme(axis.text.x=element_text(angle=70, hjust=1))
```



This graph can show us correlation between how manageable a respondents' workload is and their mental health. Those who "strongly agree" that their workload is manageable clearly report the highest levels of "excellent" mental health. Respondents who "strongly disagree" that their workload is manageable clearly reported the highest levels of "poor" mental health. Those who reported "somewhat disagree" reported the lowest levels of "excellent" mental health and second highest levels of "poor" mental health.

Critical Reflection

These visualizations reveal very interesting patterns about employment experiences and overall satisfaction in life. There were some correlations shown by the graphs that were particularly notable to me. The majority of people who reported the lowest levels of happiness work either under 20 hours per week or over 45 hours per week. Additionally, fewer work hours correlated with more loneliness, indicating that employment could be an important source of community and belonging. This being said, more analysis reveals that the quality of work is very significant in how it affects mental health and happiness. Respondents who "strongly agree" that their work is fulfilling had by far the highest proportion of "very happy" responses, while those who "strongly disagree" that their work is fulfilling clearly reported the highest proportion of "not very happy" and "not at all happy" responses. The manageability of respondents' workload also had

significant impacts on reported mental health. Those who “strongly agree” that their workload is manageable by far report the highest levels of “excellent” mental health. Those who strongly disagreed reported the highest levels of “poor” mental health. One study “The Psychosocial Quality of Work Determines Whether Employment Has Benefits for Mental Health: Results from a Longitudinal National Household Panel Survey” conducted by Butterworth, P., et al. found interesting results that support these conclusions. They report that the health benefits of being employed are reliant on the psychosocial quality of the job. The study states that “moving from unemployment into a high quality job led to improved mental health (mean change score of +3.3), however the transition from unemployment to a poor quality job was more detrimental to mental health than remaining unemployed (−5.6 vs −1.0).” We can see from both this study and my own analysis that jobs can provide a source of purpose and belonging, however the job must be fulfilling and manageable in order to have a positive impact on mental health and happiness. There are some missing aspects of this analysis that should be considered, for example I did not take into account race, sex, location, or factors such as generational wealth and income. One of my visualizations shows correlation between education level and work fulfillment. Those with a post bachelors degree reported the highest levels of fulfillment, while those with just a high school diploma most strongly disagreed that their work is fulfilling. Not everyone has the privilege to receive higher education, for a whole array of reasons. It’s important to consider these other reasons, such as poverty or difficult family situations, which can have a strong affect on mental health. This can also impact whether one has the opportunity to find fulfilling and manageable work. An area in this analysis I would want to explore further is how the pay of these jobs tie into the results. Does a fulfilling, manageable, and “high quality” job always mean a high paying job? How does income correlate with fulfillment? These are all things to consider in future research. The main conclusion to be drawn right now with the data I have is that the quality of employment strongly correlates with quality of life, mental health, and overall happiness.

Bibliography

Monnat, Shannon M., and Rhubart, Danielle C. “National Wellbeing Survey, United States, 2021.” Inter-university Consortium for Political and Social Research [distributor], 2024-06-03.

<https://doi.org/10.3886/ICPSR38879.v4>

Butterworth, P., et al. "The Psychosocial Quality of Work Determines Whether Employment Has Benefits for Mental Health: Results from a Longitudinal National Household Panel Survey." *Occupational and Environmental Medicine*, vol. 68, no. 11, 2011, pp. 806–12. JSTOR, <http://www.jstor.org/stable/23047940>.