Data types

1) Provide a URL to the dataset. # https://www.kaggle.com/osmihelp/osmi-mental-health-in-tech-survey-2017

2) Explain why you chose this dataset. I chose this dataset because I think it could provide interesting quantification of how comfortable people in tech feel discussing and seeking help for their mental health conditions since it is about 700 rows and 123 colums.

3) What are the entities in this dataset? How many are there? The entities are survey responses. Each row represents one person's survey response.

#_4) How many attributes are there in this dataset? There are 123 attributes in this dataset.

5) What is the datatype of each attribute (categorical -ordered or unordered-, numeric -discrete or continuous-, datetime, geolocation, other)? Write a short sentence stating how you determined the type of each attribute. Do this for at least 5 attributes, if your dataset contains more than 10 attributes, choose 10 of them to describe.

Num	Name	Type	Description
1	Are you self-employed?	Categorical Attribute (discrete, unordered)	I could see that this attribute takes the discrete set of the binary numeral system; 1 represents self-employed, 0 represents not self-employed
2	What country do you live in?	Categorical Attribute (discrete, unordered)	Countries is a finite set that is discrete so this is categorical
3	Does your employer provide mental health benefits as part of healthcare coverage?	Categorial Attribute (discrete, unordered)	The data in this column/aatribute contains the response "Yes", "No", "I don't know", or "Not eligible", which is a distinct, discrete set of responses. It is unordered because there is no distance or order between yes, no, or "I don't know".
4	Describe the conversation you had with your employer about your mental health, including their reactions and what actions were taken to address your mental health issue/questions.	Unstructured Response Text Attribute	The responses are sentences/strings that users have typed in, describing a conversation.
5	Was your anonymity protected if you chose to take advantage of mental health or substance abuse treatment resources with previous employers?	Ordinal Attribute (discrete)	There are 4 distinct responses: "yes, always", "no", "sometimes", and "I don't know" which are ordered by most to least protection.

Num	Name	Type	Description
6	Overall, how much importance did your previous employer place on mental health?	Categorical Attribute (ordered, numerical, discrete)	These responses took a distinct numerical rating from 0-10
7	If a mental health issue prompted you to request a medical leave from work, how easy or difficult would it be to ask for that leave?	Ordinal Attribute (discrete)	The set of responses range from "very easy" to "neither difficult nor easy" to "difficult" or "very difficult", so this attribute is a ranking and ordered from most difficult to most easy.
8	Do you know the options for mental health care available under your employer-provided health coverage?	Categorical Attribute (discrete, unordered)	This attribute takes on the responses of "yes" or "no" so it is discrete set of possible responses/data.
9	What is your race?	Categorical Data (unordered)	The races are distinct groups that do not have a particular order.
10	What is your age?	Numerical (discrete)	The age is measured in years so it is discrete.

6) Write R code that loads the dataset using function `read_csv`*. Were you able to load the data successfully? If no, why not?* Yes, I was able to load it.

```
library(tidyverse)
# loading code goes here
dataset <- read_csv("OSMIMentalHealthinTechSurvey2017.csv")
```

```
## Warning: Duplicated column names deduplicated: 'Describe the conversation your
## coworker had with you about their mental health (please do not use names).'
## => 'Describe the conversation your coworker had with you about their mental
## health (please do not use names)._1' [46], 'Anxiety Disorder (Generalized,
## Social, Phobia, etc)' => 'Anxiety Disorder (Generalized, Social, Phobia,
## etc)_1' [64], 'Mood Disorder (Depression, Bipolar Disorder, etc)' => 'Mood
## Disorder (Depression, Bipolar Disorder, etc)_1' [65], 'Psychotic Disorder
## (Schizophrenia, Schizoaffective, etc)' => 'Psychotic Disorder (Schizophrenia,
## Schizoaffective, etc)_1' [66], 'Eating Disorder (Anorexia, Bulimia, etc)'
## => 'Eating Disorder (Anorexia, Bulimia, etc)_1' [67], 'Attention Deficit
## Hyperactivity Disorder' => 'Attention Deficit Hyperactivity Disorder_1' [68],
## 'Personality Disorder (Borderline, Antisocial, Paranoid, etc)' => 'Personality
## Disorder (Borderline, Antisocial, Paranoid, etc)_1' [69], 'Obsessive-Compulsive
## Disorder' => 'Obsessive-Compulsive Disorder_1' [70], 'Stress Response Syndromes'
## => 'Stress Response Syndromes_1' [72], 'Dissociative Disorder' => 'Dissociative
## Disorder_1' [73], 'Substance Use Disorder' => 'Substance Use Disorder_1' [74],
## 'Addictive Disorder' => 'Addictive Disorder_1' [75], 'Other' => 'Other_1' [76],
## 'Anxiety Disorder (Generalized, Social, Phobia, etc)' => 'Anxiety Disorder
## (Generalized, Social, Phobia, etc)_2' [77], 'Mood Disorder (Depression, Bipolar
## Disorder, etc)' => 'Mood Disorder (Depression, Bipolar Disorder, etc)_2' [78],
```

```
## 'Psychotic Disorder (Schizophrenia, Schizoaffective, etc)' => 'Psychotic
## Disorder (Schizophrenia, Schizoaffective, etc)_2' [79], 'Eating Disorder
## (Anorexia, Bulimia, etc)' => 'Eating Disorder (Anorexia, Bulimia, etc)_2' [80],
## 'Attention Deficit Hyperactivity Disorder' => 'Attention Deficit Hyperactivity
## Disorder_2' [81], 'Personality Disorder (Borderline, Antisocial, Paranoid,
## etc)' => 'Personality Disorder (Borderline, Antisocial, Paranoid, etc)_2' [82],
## 'Obsessive-Compulsive Disorder' => 'Obsessive-Compulsive Disorder_2' [83],
## 'Post-traumatic Stress Disorder' => 'Post-traumatic Stress Disorder_1' [84],
## 'Stress Response Syndromes' => 'Stress Response Syndromes_2' [85], 'Dissociative
## Disorder' => 'Dissociative Disorder_2' [86], 'Substance Use Disorder'
## => 'Substance Use Disorder_2' [87], 'Addictive Disorder' => 'Addictive
## Disorder_2' [88], 'Other' => 'Other_2' [89], 'Why or why not?' => 'Why or why
## not?_1' [100], 'Other' => 'Other_3' [118]
```

```
## Wrangling

#1) My pipeline computes the mean age of those who completed the survey
#and the mean rating responders gave for the question
#"How much importance #does your employer place on mental
#health" which gives us an idea of
#how generally employees feel
#about the support their tech employers
#provide for mental health issues.

#```{r pipeline}
# pipeline goes here
select(dataset, 'Overall, how much importance does your employer place on mental health?',
       'Is your employer primarily a tech company/organization?',
       'What is your age?')
```

```
## # A tibble: 756 x 3
##    `Overall, how much importance d~ `Is your employer primaril~ `What is your a~
##                               <dbl>                       <dbl>            <dbl>
##  1                                0                           1               27
##  2                                2                           1               31
##  3                                1                           1               36
##  4                                5                           1               22
##  5                               NA                          NA               52
##  6                                5                           1               30
##  7                               10                           1               36
##  8                                8                           1               38
##  9                                7                           0               35
## 10                               NA                          NA               36
## # ... with 746 more rows
```

```
# groups mean
dataset %>%
 group_by(is_a_tech_company_0_or_1=as.numeric(dataset$`Is your employer primarily a tech company/organi
  summarize(importance_employee_feels_is_placed_on_mental_health_rating=mean(as.numeric(`Overall, how mu
            importancePhys=mean(as.numeric(`Overall, how much importance does your employer place on phy


## Plotting

#1) This plot shows how much importance survey responders feel
```

```
  #their employer places on mental health at tech (1) or
  #non-tech (0) companies.

#```{r plot}
# plot goes here
ggplot(dataset,
       mapping=aes(x=is_a_tech_company_0_or_1,
                   y=importance_employee_feels_is_placed_on_mental_health_rating)) + geom_bar(stat = "id
```

```
## Warning: Removed 1 rows containing missing values (position_stack).
```