

1. Change the file to tab separated.txt file
2. `cat mental-heath-in-tech-2016_20161114.tsv | egrep '^0|^A.*'> MHT.tsv # to keep the non-self-employed.`
3. `cat MHT.tsv | cut -f 1-5, 7, 9- 16, 37, 39, 41-48, 51, 53, 54, 56-59, 63 > MHT_columns.tsv # to select the columns of interest.`
4. `cat MHT_columns.tsv | sed 's/\t/angela/g' | sed 's/angelaangela/angelanot applicableangela/g' | sed 's/angela/\t/g' > MHT_columns2.tsv # to clean the extra tab and white space in the file.`
5. `cat MHT_columns2.tsv | sed 's/26-100/ss/' | sed 's/100-500/sm/' | sed 's/500-1000/mm/' | sed 's/More than 1000/ml/' | sed 's/43470.00/ll/' | sed 's/43641.00/ll/' > MHT_columns3.tsv # change the company size to categorical labels`
6. `cat MHT_columns3.tsv | cut -f 2-32 > MHT_columns4.tsv # remove the first columns since the value all equals to 0`
7. `cat MHT_columns4.tsv | sed 's/(..\t[0-9]\t)not applicable/\11/' > MHT_columns5.tsv # clean out all the blanks for "Is your primary role within your company related to tech/IT?" column. The ones have blank are all from the IT company.`
8. `cat MHT_columns5.tsv | sed 's/(..\t[0-9]\t[0-9]\t)Not eligible for coverage \N\A/\1No/' > MHT_columns6.tsv # change the "not eligible for coverage" to "No"`
9. `cat MHT_columns6.tsv | sed '1141d' | sed '1128d' | sed '1019d' | sed '879d' | sed '866d' | sed '861d' | sed '752d' | sed '719d' | sed '517d' | sed '515d' | sed '498d' | sed '447d' | sed '300d' | sed '135d' | sed '79d'> MHT_columns7.tsv #removed the rows that doesn't have complete data`
10. `cat MHT_columns7.tsv | cut -f1-15 > MHT_8.tsv`
11. `cat MHT_columns7.tsv | cut -f16 > cutf16.tsv`
12. `cat cutf16.tsv | sed 's/"No, I don't think it would"/No/" | sed 's/"No, it has not"/No/" | sed 's/"Yes, I think it would"/Yes/" | sed 's/"Yes, it has"/Yes/' > cutf16_cleaned.tsv #Generalize the answers`
13. `paste MHT_8.tsv cutf16_cleaned.tsv > 1_16.tsv #Merge the two files to one file that has columns 1 to 16.`
14. `cat MHT_columns7.tsv | cut -f17 | sed 's/"No, I don't think they would"/No/" | sed 's/"No, they do not"/No/" | sed 's/"Yes, I think they would"/Yes/" | sed 's/"Yes, they do"/Yes/'> cutf17_cleaned.tsv #Generalize the answers`
15. `paste 1_16.tsv cutf17_cleaned.tsv > 1_17.tsv #append column 17 to the 1_16.tsv file`

16. `cat MHT_columns7.tsv | cut -f 18-27 > 18_27.tsv`

17. `paste 1_17.tsv 18_27.tsv > 1_27.tsv`

18. `MHT_columns7.tsv | cut -f 28 | sort | uniq -c`

#below screenshot displays a portion of the output from “MHT\_columns7.tsv | cut -f 28 | sort | uniq -c”

```
  2 Agender
  1 Androgynous
  1 Bigender
  1 cisdude
  1 Cis female
  1 Cisgender Female
  1 cis male
  1 Cis male
  1 Cis Male
  1 cis man
  1 Cis-woman
  1 Dude
  1 Enby
 20 f
 24 F
   1 fem
 77 female
   2 female
121 Female
   1 Female
   6 Female
   1 Female assigned at birth
   1 female-bodied; no feelings about gender
   1 Female or Multi-Gender Femme
   1 "Female (props for making this a freeform field, though)"
   1 female/woman
   1 fm
   1 Genderfluid
   1 Genderfluid (born female)
   1 genderqueer
   1 Genderqueer
```

19. `cat MHT_columns7.tsv | cut -f 28 > 28.tsv`

20. `cat 28.tsv | sed 's/Genderfluid/na/' | sed 's/Enby/na/' | sed 's/genderqueer/na/' | sed 's/Genderqueer/na/' | sed 's/N/A/na/' | sed 's/nb masculine/na/' | sed 's/non-binary/na/' | sed 's/none of your business/na/' | sed 's/not applicable/na/' | sed 's/Other/na/' | sed 's/Queer/na/' > 28_4.tsv # change all non-male or non-female to na`

21. `cat 28_4.tsv | sed 's/\"male 9:1 female, roughly\"/M/' | sed 's/\"Male (trans, FtM)\"/M/' | sed 's/. *F. */F/' | sed 's/. *f. */F/' | sed 's/. *w. */F/' | sed 's/. *W. */F/' | sed 's/. *m. */M/' | sed 's/. *M. */M/' | sed 's/Non. */na/' | sed 's/. *ud. */M/' > 28_cleaned.tsv # clean the columns to only F/M/na`
22. `paste 1_27.tsv 28_cleaned.tsv > 1_28.tsv`
23. `cat MHT_columns7.tsv | cut -f 29-31 > 29_31.tsv`
24. `paste 1_28.tsv 29_31.tsv > MHT_9.tsv`
25. `MHT_9.tsv | egrep 'United States of America | age' > MHT_10.tsv #select only the US workers`
26. `cat MHT_10.tsv | egrep -v '323' > MHT_11.tsv # get rid of the line that has age 323`
27. `cat MHT_11.tsv | cut -f 1-28 > 1_28_2.tsv`
28. `cat MHT_11.tsv | cut -f 30-31 > 30_31.tsv`
29. `paste 1_28_2.tsv 30_31.tsv > MHT_final.tsv #cut off column 29 since the values are the same`
30. `cat MHT_final.tsv | sed 's/N\A/NA/g' > MHT_final2.tsv`