FIT5147 DATA EXPLORATION AND VISUALIZATION

Data Exploration Project Report

State of global happiness and the potential factors affecting it

Tutor

Farah Tasnuba Kabir

Tutorial Number 04-P1

Submitted by

Manali Prakash Choudhary

30151198

Table of Contents

Table of Figures

1. Introduction

Humanity has come a long way thinking about happiness and developing various ways of acquiring it. Everyone is striving and struggling to accomplish goals or is performing certain actions to achieve happiness for oneself or for others. Hence, happiness is the fundamental desire of mankind. But happiness can be recognized as an emotion and as a measure to judge the quality of life as in whole. But can we really measure happiness? This can be debatable as the measures of happiness are different for different cultures and people across the globe. Some say money is the new definition for happiness, some say family is. Some rank the developed countries as the happiest and some consider mental health and peace to be important. In various countries mental health conditions are still not taken seriously. As people spend most of their time in workplace, under mental stress and constant pressure, awareness and frequency of mental disorder should be considered important contributing to the happiness indicators. Hence, all these measures contribute to the subjective wellbeing of human life (Hall & Helliwell, 2014). Behavior of different regions on different measures can help add value to the quality of life of people and review the progress of nations in right direction. Building a standard fair platform or measures and such study of happiness factors in all countries around the globe can help the policy makers, governments and communities to come up with better solutions and policies for the existing issues to improve the life of their citizens.

So, can we really find any common factors to judge the happiness of people from different countries around the globe? How can then such happiness factors contribute to improve the quality of life of mankind. These questions contributed to the motivation of performing a detailed data exploration on the *state of global happiness and the potential factors affecting it*. This data exploration will help answer the following questions on the six subjective wellbeing factors - economic production, social/family support, health/life expectancy, freedom, trust in government/absence of corruption, and generosity and also mental health,

a) What is the behavior of all the countries/ regions according to the happiness score and the top subjective wellbeing factors most affecting the happiness score for all the years?
b) What are the trends of the respondents of the mental health survey in order to explore the potential causes?
c) What is the relation of the happiness rank of countries and the employer support available for mental health wellbeing in that country?

2. Data Wrangling

A) The World Happiness Report Statistical Data for year 2015, 2016 and 2017

The World Happiness Report is a publication of the Sustainable Development Solutions Network, powered by data from the Gallup World Poll for around 157 countries and is released by United Nations (UN) every year since 2012, with the aim of helping the world governments and policy makers to enable such policies benefiting the quality of human development. This data basically contains the happiness scores of different countries and their rankings, based on the answers acquired to the main life evaluation questions on subjective wellbeing raised in the poll. The happiness score ranges from 0-10 and estimates the dependencies of the score on the six factors that potentially contributes to the happiness index of the countries. These scores are evaluated in comparison to the scores of an imaginary country "Dystopia", that has world's least-happy people or least possible scores for those six factors. The idea of having "Dystopia" is to have standard set against all countries for favorable comparison and positive values for each key factor. The links for the three data sources for year 2015, 2016 and 2017 respectively are as below,

(https://worldhappiness.report/ed/2015/),(https://worldhappiness.report/ed/2016/), (https://worldhappiness.report/ed/2017/)

The Data Wrangling process for this dataset has been carried out using *R studio*. The following steps are performed for the same.

1. Imported all the three datasets in R studio and analyzed them according to the research questions.

2. Since there are three separate datasets for year 2015, 2016 and 2017 respectively, they need to be combined vertically into a single dataset for easy operations and exploration. Hence, added a new column "Year" in all three datasets with the value of the respected dataset year.

3. An important step for vertical merging of datasets is assuring the datasets have same columns. Hence, removed unwanted and uncommon columns. 'Region' column with certain values was missing in the '2017' dataset, which was added before combining based on the values from 2015 and 2016 dataset.

4. Performed vertical join of the three datasets using "rbind" in R and gave appropriate names to the columns in the combined dataset.

5. The numerical values in the dataset possessed values with up to 7 decimal places. Hence, rounded the numerical values to two decimal places.

6. Imported the combined dataset into Tableau Desktop for further analysis and data checking.

Handling '0' values

The combined dataset for Happiness scores showed data value '0' in many fields. But here, data with value '0' is numerical data, i.e. score given for the six happiness factors range from 0 – 10 and are corelated with each other contributing to the final happiness score in this dataset. Hence, they have value and so were not imputed. A glimpse of the dataset having numerical value ranging from 0 is shown in the figure below for summary statistics.

```
Happiness.Rank   Happiness.Score Economy.GDP.per.Capita
Min.   :  1.00   Min.   :2.690   Min.   :0.0000
1st Qu.: 40.00   1st Qu.:4.510   1st Qu.:0.6025
Median : 79.00   Median :5.285   Median :1.0000
Mean   : 78.83   Mean   :5.371   Mean   :0.9277
3rd Qu.:118.00   3rd Qu.:6.235   3rd Qu.:1.2500
Max.   :158.00   Max.   :7.590   Max.   :1.8700
    Family       Health.Life.Expectancy    Freedom
Min.   :0.0000   Min.   :0.0000         Min.   :0.000
1st Qu.:0.7900   1st Qu.:0.4000         1st Qu.:0.300
Median :1.0300   Median :0.6300         Median :0.420
Mean   :0.9902   Mean   :0.5802         Mean   :0.403
3rd Qu.:1.2300   3rd Qu.:0.7700         3rd Qu.:0.520
Max.   :1.6100   Max.   :1.0300         Max.   :0.670
Trust.Government.Corruption   Generosity
Min.   :0.0000              Min.   :0.0000
1st Qu.:0.0600              1st Qu.:0.1500
Median :0.1000              Median :0.2200
Mean   :0.1349              Mean   :0.2422
3rd Qu.:0.1700              3rd Qu.:0.3200
Max.   :0.5500              Max.   :0.8400
> |
```

Figure 1. Summary statistics of the Happiness Report Data

B) Data from Survey on Mental Health

Data from Survey on Mental Health that measures the attitude of mental health and frequency of mental health disorders in the Tech Workplace in 2014-15 from all over the globe, acquired from Open Sourcing Mental Illness under CC BY-SA 4.0. This dataset is based on the responses acquired from the respondents where each column states each question asked regarding the workplace attitude towards mental health and frequency of mental health disorders. The link for the data source is given below.

(https://osmihelp.org/research)

The Data Wrangling process for this dataset has been carried out using *R studio*. The following steps are performed for the same.

1. Imported all the dataset in R studio and analyzed them according to the research questions.

2. Removed unwanted columns which are not useful and affecting our research.

3. Changed the unstandardized format of the "Timestamp" column to a standard date format and extracted year, making it a "Year" column.

4. The "Gender" column possessed various irregular gender values, this being a survey response data. Hence, handled this by performing regex operations in R. The before and after images of the column are as shown.

5. Imported the wrangled dataset into Tableau Desktop for further analysis and data checking.

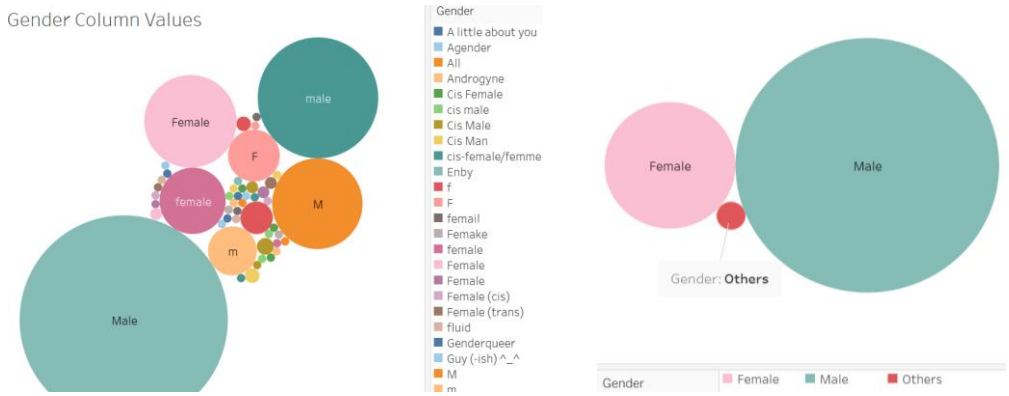


Figure 2. Gender column before and after performing regex

Handling 'NA' values

Handling 'NA' values in categorical textual data is thoroughly depended on your dataset behavior and meaning of the variables and categories. Hence, such suggested methods according to a study in a blog post on *ResearchGate* (Nwegbu, 2015) were helpful in our case. The 'self_employed' column possessed 'NA' values contributing to just 1.43% of the total data, hence, these values were imputed with the most frequent values for the column i.e. No. The same is as shown in Fig. 3.

Figure 3. Bar plots showing the 'NA' value handling for 'self_employed' column

'NA' values in column 'work_interfere' were imputed with a fair standardized category 'Don't know', which is in uniformity with categories of other columns in the dataset. The dataset is not large enough and also the rows with 'NA' contribute to around 21% of the total data. The other responses of such respondents are important to be recorded from this research point and removing them may risk in affecting the overall research. Also, since this is a survey response data, manipulating or imputing 'NA' data with any other potential value categories can affect the validation and reliability of the exploration outcomes obtained through it.

3. Data Checking

The Data checking process for this project has been carried out using *Tableau Public* and *R studio*.

A) The World Happiness Report Statistical Data for year 2015, 2016 and 2017.

The happiness score statistical data table includes subjective wellbeing factors which contribute to the overall happiness score of a country. Finding a relation between them and on the happiness score is vital in our study. As suggested by the studies (Bock, 2020), a correlation matrix shows a correlation between the variables by the correlation variables and works as a diagnostic for more advanced analysis. Handling of missing or '0' values is an important factor to be addressed before plotting a correlation matrix. Hence, after doing so in the data wrangling step, we can now move ahead with the plotting of correlation matrix. The correlation of the subjective wellbeing factors on each other and on the happiness score/ rank was performed in R as shown in the below figure for further detailed exploration and analysis of the data.

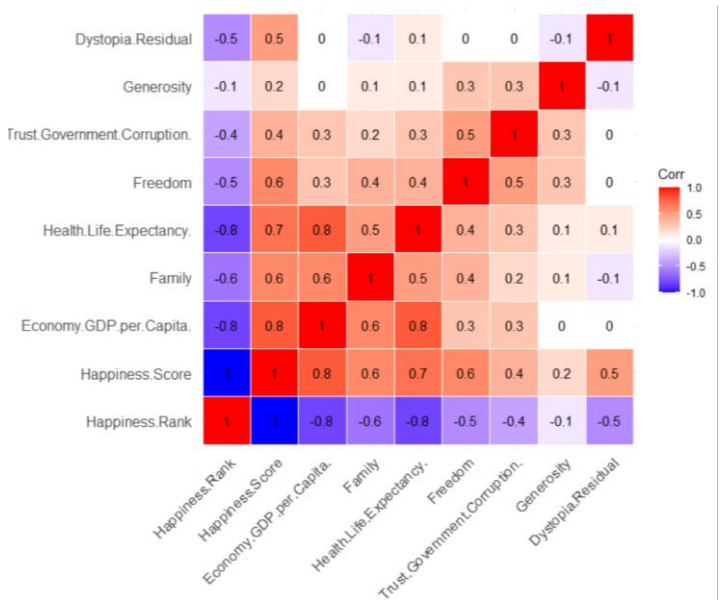


Figure 4. Correlation Matrix of the Happiness Report Data

B) Data from Survey on Mental Health

The 'Age' column in this dataset showed outliers and jargon values. Hence, the data checking implemented using boxplot for this column was ineffective due to the high range values of the outliers. Exploration obtained through a histogram (left skewed) gave a clear idea of the outliers and their count as well as shown. The outliers were removed in R by limiting the age values to a minimum and maximum range of Age and visualized in Tableau as shown in Fig. 5. Using other outlier removal techniques like n-sigma,

interquartile range proved inefficient in giving the logical values, as logically exploring, the age of the survey respondents should be in working or retired class.

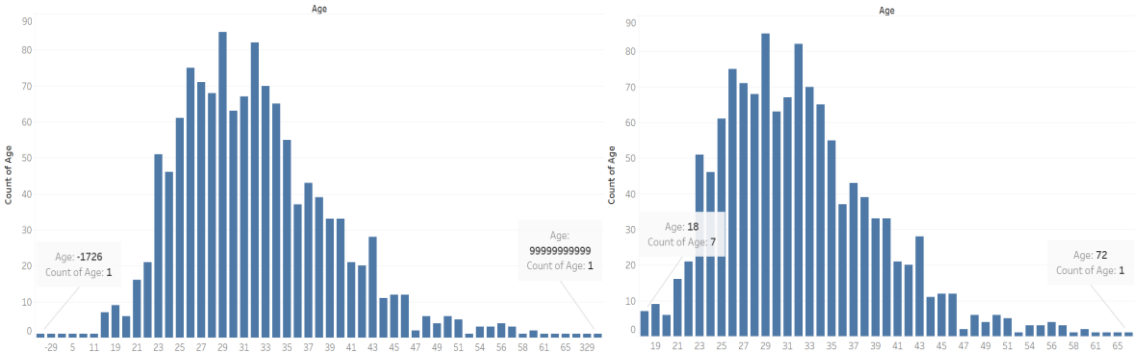


Figure 5. Histogram with and without the outliers for 'Age'

```
>   summary(Age)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  18.00   27.00   31.00   32.08   36.00   72.00
```

Figure 6. Summary statistics for 'Age' after removing the outliers

C) Merged dataset from World Happiness Report and Mental Health Survey

In order to answer the question c, a merged dataset is created using an inner join. The question demands a column showing the percentage of categorical responses from the total responses from the country for all the countries. Hence a column named 'Percent' is calculated and added in this dataset.

4. Data Exploration

The Data Exploration and Visualization process for this project has been carried out using *Tableau Public*.

a) What is the behavior of all the countries/ regions according to the happiness score and the top subjective wellbeing factors most affecting the happiness score for all the years?

We have already figured out the correlation of the factors for Happiness score while we performed the exploration for the data checking using the correlation matrix as shown earlier. The top three subjective wellbeing factors affecting the Happiness score are Economy, Health and Freedom. The following three figures in Fig. 6 show the Bubble plot for the detailed low-level exploration for the years 2015, 2016 and 2017, respectively. Australia and New Zealand, Western Europe, Latin America and Caribbean and North America are the regions who are above average for all the factors and happiness score high for all the three years, while, Southern Asia and Sub–Saharan Africa show below average scores. Over the period, there is an improvement in the state of economy all around the world, including the Sub-Saharan African region. There are some fluctuations towards the average axis for the state of freedom for all the regions from 2015 to 2017, specially for the Sub-Saharan and Middle- east and Northern African region, the accumulation can be seen nearing to the average freedom score over the period. For the Health factor, Southern Asia and Sub-Saharan Africa show low scores throughout the period, poverty being one of the causes. In fact, over the period the regions having above average health status in 2015 are accumulating near the average score or even lower over the period.

Distribution of Happiness Score around the globe in relation to the Economy, Health and Freedom (2015,2016 & 2017)

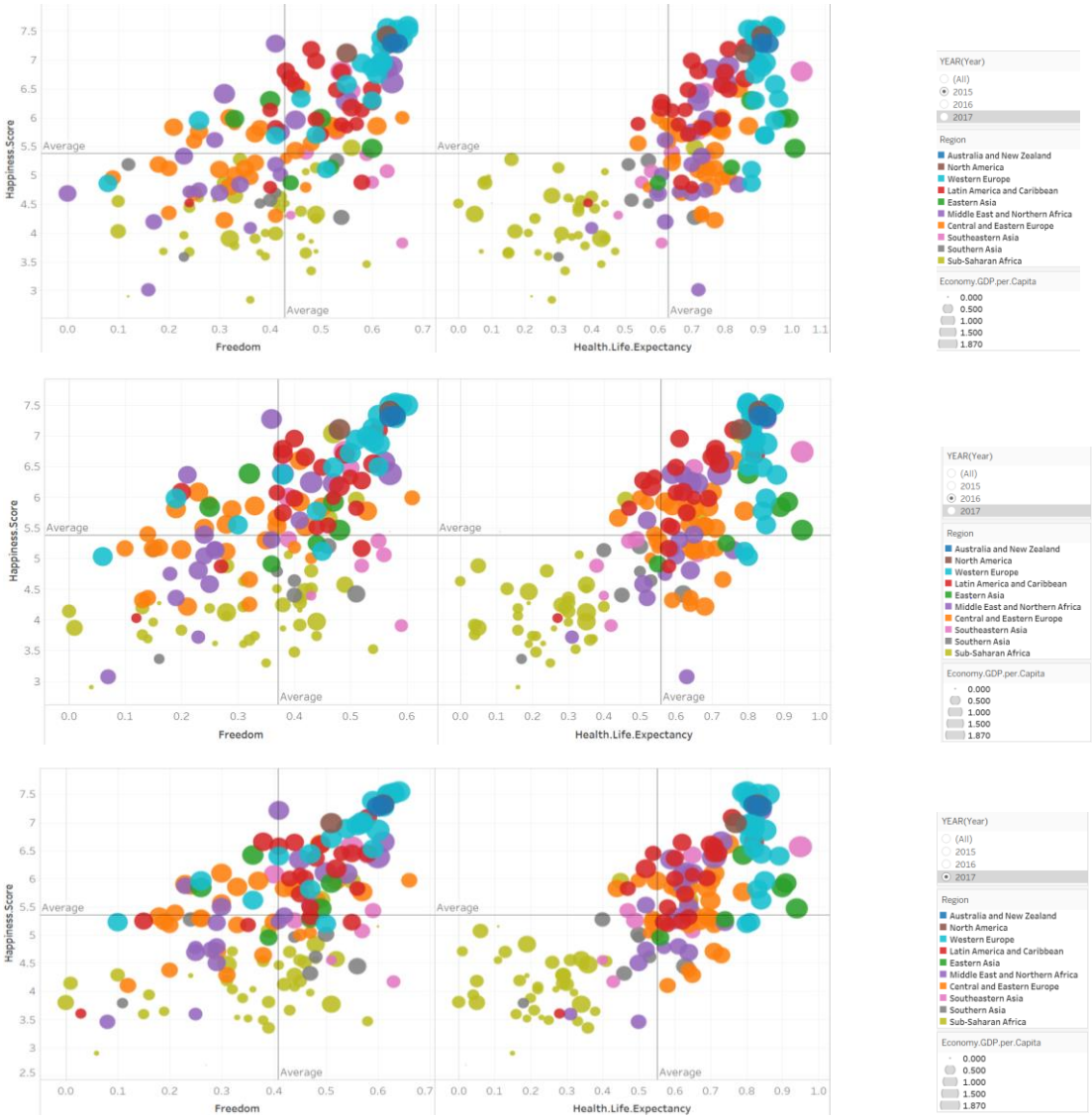


Figure 7. Distribution of countries/ region according to economy, health, freedom and happiness score

Hence, it can be concluded that though the status of the economies has improved, it is not applicable to the Freedom and Health status throughout the world. Also, some regions of the world are suffering with lower scores of the subjective wellbeing factors contributing to lower Happiness score. This might be because, the subjective wellbeing factors like health, economy etc. are related with the major causes like poverty, unemployment, high population, cultural issues which can take decades to eradicate. The bar charts of the distribution of Health, Economy, Freedom and Happiness score for all the years as plotted below (Fig. 7) which depict the summary of the bubble charts and hence, supports our conclusion as well. One surprising observation to note here is that the average happiness score is almost constant for all the three years. This might be because, as we have seen, though the economy is improving over the period, it is not the same case for other factors like health, freedom etc.

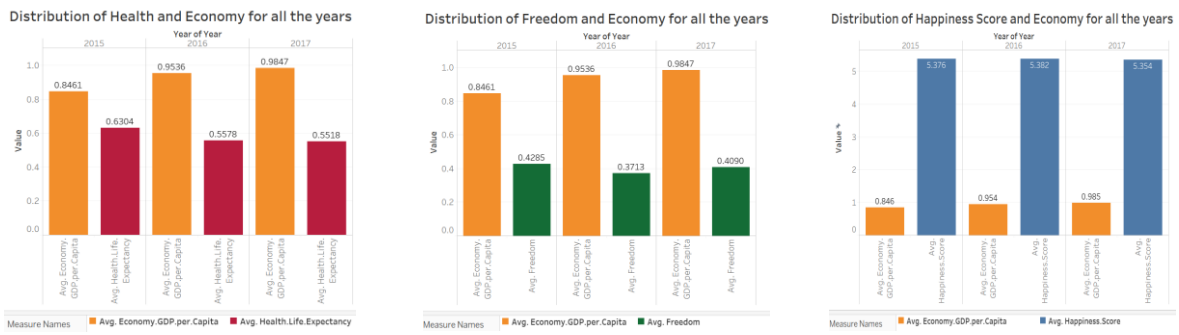
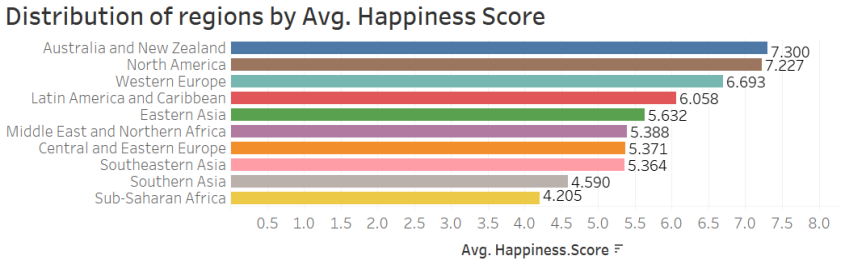


Figure 8. Distribution of Health, Economy, Freedom and Happiness score for all the years

The top 10 countries and the regions based on their happiness score are also plotted in the Fig. 8. Among the top 10 countries, we can see that majority of the countries belong to Western Europe region. But, region wise, Australia and New Zealand, Western Europe, Latin America and Caribbean and North America are the regions with highest happiness score.

Hence, this study shows that happiness score depends on the subjective wellbeing factors chosen for this study. Countries with high scores for these factors automatically shows high happiness scores

.



Selecting by Avg_Happiness_Score
A tibble: 10 x 2

	Country	Avg_Happiness_Score
	<fct>	<dbl>
1	Switzerland	7.53
2	Denmark	7.53
3	Iceland	7.52
4	Norway	7.52
5	Finland	7.43
6	Canada	7.38
7	Netherlands	7.37
8	Sweden	7.31
9	New Zealand	7.31
10	Australia	7.29

Figure 9. Regions and Top 10 countries by Avg. Happiness Score

b) What are the trends of the respondents of the mental health survey in order to explore the potential causes?

Let us have a look at the trends of the respondents of the mental health survey. Here we start with the Age and Gender factor. By the exploration of this data set while data checking for the columns Gender and Age it was evident that Males has contributed to the majority responses of this survey and the median Age of respondents is around 32, with the range 18 – 72. The *Treatment* column here is the most important field in this dataset and refers to the question '*Have you sought treatment for a mental health condition?*' in the survey. As shown in Fig. 9 majority males have responded as Yes to this question, with the box plot showing values evenly distributed. The box plot for male and Female show equal number of responses for both Yes and no category. But, for Others, majority respondents have responded Yes to the question suggesting that almost all of them have suffered mental health issues where social/ cultural factors can be one important issue.

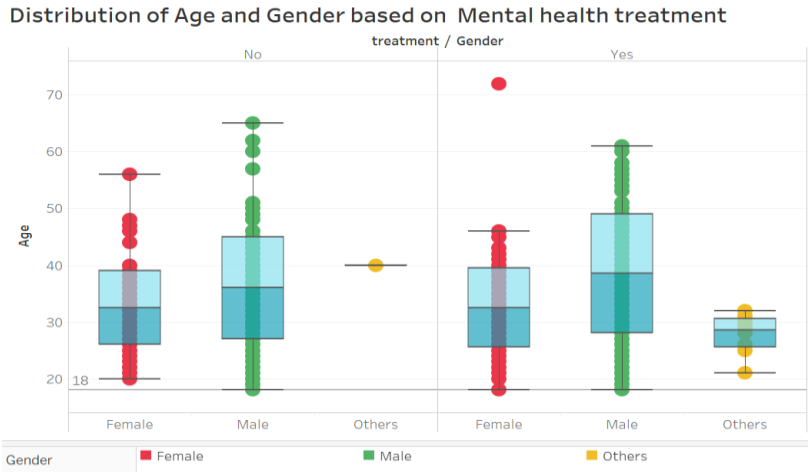


Figure 10. Distribution of Age and Gender based on Mental health treatment

Next, we check for the trends to analyse the effects of mental health issues on work routine. The columns *leave*, *work_interfere* in this dataset refers to the *questions How easy is it for you to take medical leave for a mental health condition?* and *If you have a mental health condition, do you feel that it interferes with your work?* respectively. The below Fig. 10 is a balloon plot which is a variation of the bar plot. The high responses of Don't know for treatment 'No' in leave and work interfere is due to the imputation of Don't know value for NA values, which is pointless for our study. We have high number of responses saying that mental health illness has affected their work Sometimes and Often where taking leave was Somewhat difficult, Somewhat easy and very easy.

There are also majority respondents saying 'Don't know' for the leave factor suggesting such respondents have hardly applied leave for their illness or haven't acknowledged it though it has stiff affected their work.

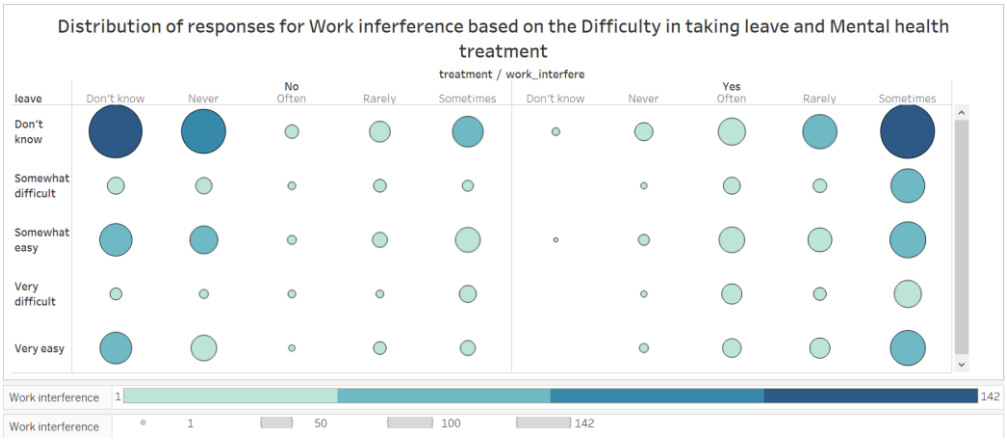


Figure 11. Distribution of count of responses for work interference vs leave and treatment

Let us now analyse the potential causes or factors for the mental health illness among our respondents. A simple bar plot as shown below helps us answer this question with the fields referring to the following questions,

- *self_employed - Are you self-employed?*
- *family_history - Do you have a family history of mental illness?*
- *tech_company - Is your employer primarily a tech company/organization?*

Respondents who have family history of mental health illness, are not self – employed and are working for a tech company responded with the highest count for undergone treatment of mental illness. Hence, according to the survey this suggests that not being self-employed, working in a tech company and having a family history of mental health illness can be potential factors contributing to the mental health issues in work environment.

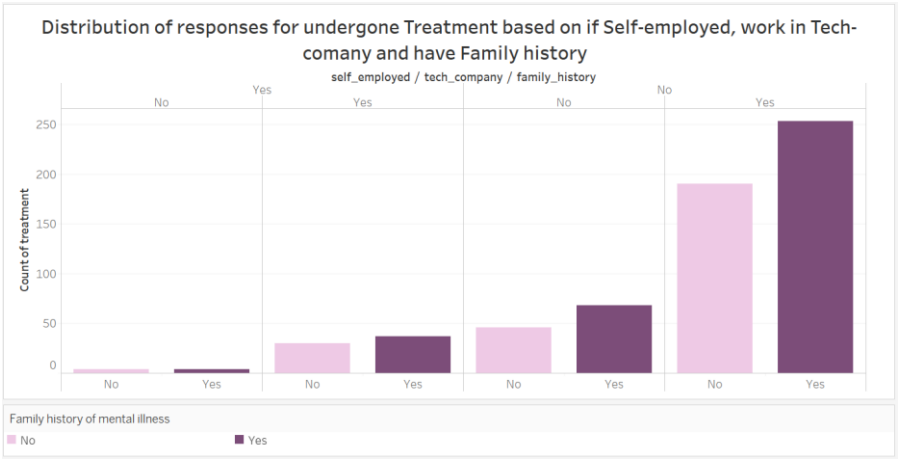


Figure 12. Distribution of count of responses for treatment vs self-employment, tech-company & family history

c) What is the relation of the happiness rank of countries and the employer support available for mental health wellbeing in that country?

In order to answer this question, we need to join our datasets for mental illness and happiness report. The same is done in Tableau Desktop by joining with an inner join by 'Country' column. The column seek_help from mental health dataset is found useful for this exploration which refers to the question *Does your employer provide resources/ support about mental health issues and how to seek help?* for the survey. We have already added a column 'Percent' during data checking which allows us to know the percentage of categorical responses from the total responses from the country for all the countries. Hence, we use this column to plot the pie chart depicting these responses against the happiness rank of the country using a map. As seen, majority responses are from United States followed by United Kingdom and Canada. The proportional symbol map plotted in the below Fig. 12, shows that the numbers

of positive responses are surprisingly low for all the countries. Majority responses from majority countries are negative for this question. Majority countries/ regions with high happiness rank, namely, Northern America, Western Europe, Australia and New Zealand show some percent positive responses for the question. Australia shows the maximum percent of positive responses among all the countries

Hence, it can be concluded that the world still needs to show a heavy improvement for the condition of mental health support and wellbeing and receive the same level of importance as the physical health. Also, as countries with high happiness rank responded better as compared to those with low happiness rank, we can conclude that mental health wellbeing and employer support does have a dependent relation on each other.

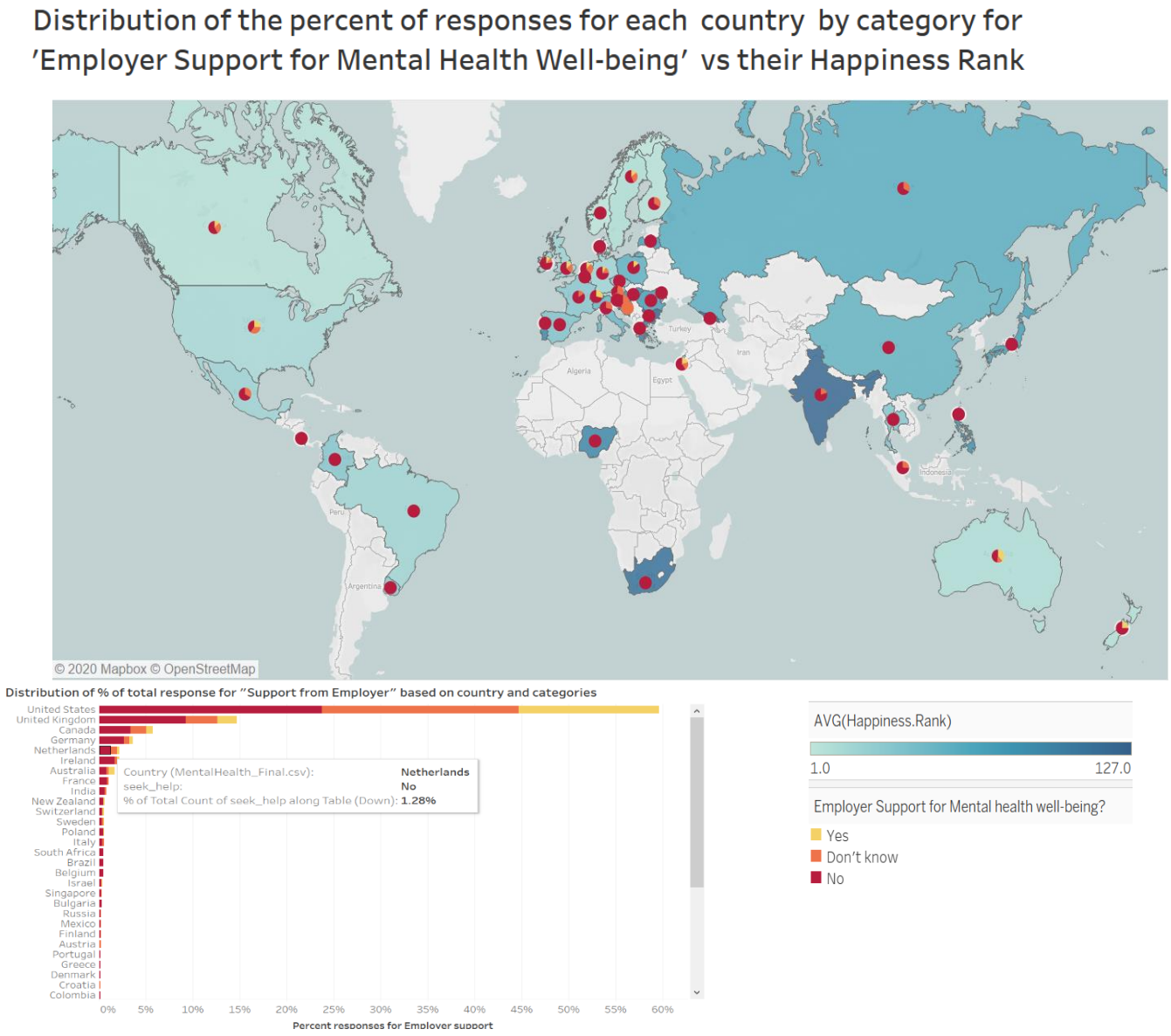


Figure 13. Distribution of responses for 'Support from Employer' for each country vs their Happiness Rank

5. Conclusion

The Happiness State Report Statistical dataset along with the Mental Health Survey dataset gave multiple detailed findings and conclusions. Starting from a curiosity to find the relevance of Happiness Score and factors led to some surprising, happy and hopeful conclusions. Countries with high happiness score have high scores in the subjective wellbeing scores and also the employer support for mental wellbeing. Countries with low scores are struggling hard to rise above the average. As discussed earlier, this might be because, the subjective wellbeing factors like health, economy etc. are related with the major causes like poverty, unemployment, high population, cultural issues which can take decades to eradicate. Factors of subjective wellbeing significantly contribute to the happiness of a country, but also does the mental health wellbeing. Tech company, family history of mental illness and self-employment are some of the important factors contributing to the causes of Mental Health illness among the survey respondents. Work environment may affect majorly to the mental health condition. A support from the employer in such case can be very fruitful in contributing to the mental health wellbeing and hence, the happiness state and happiness score of a country.

6. Reflection

The data exploration research study for the happiness state and the mental health state of the world was carried using standard data exploration process. This included steps like data cleaning/ reformatting, data checking, data exploration and visualization steps where each step is equally valuable to tell the true data story. The data cleaning and reformatting allowed us to explore R studio for multiple functions, like ways to join tables, i.e. as we have implemented vertical table join. Also, implemented Regular expression operations to correct the column values and implemented null values handling. In our study, as the datasets were numerical and categorical with textual data, the null values handling, and exploration was different for both. In terms of the numerical dataset, the values '0' contributed to the final happiness score column and hence could not be imputed. Data checking generally consists the process to remove the outliers and checking the behavior of the data for further detailed analysis. The most classic way of outlier detection is boxplot, but here, we used histogram to do so. Hence, we learned that the most classic methods are not always suitable in every study. The same results can be yielded by using visualizations more suitable according to our dataset. This exploration led to the plotting of correlation matrix, which made the correlation very clear and mapped the path for further detailed low-level study. For the categorical textual dataset, such a correlational matrix is not possible. Hence, that needed to be understood using simple bar charts and the purpose of the fields and values logically. Hence, here we get to play with numerical and textual(categorical) datasets to realize that exploring them implements different methods. In both cases the statistical tests prove important to justify the approaches and give the data exploration the correct direction. The data exploration and visualization process included modern versions of the classic plots/visualization depending on whether the data is textual or numerical. The exploration steps during the data checking process allowed some detailed low-level data visualizations. Here, we have majorly studied and plotted multivariate data visualizations. Sometimes, the multivariate data visualizations can be too low-level making it difficult and time consuming to come up to a conclusion. In this case, a supporting high-level visualization can be used for better understanding. The same idea was implemented, which further added more value to this project.

In hindsight, addition of more datasets to analyse multiple factors (e.g. social/ cultural) affecting the happiness score and mental health wellbeing could be interesting. Further exploration in handling different types of data is also fascinating. As both the datasets were survey based, it would have been great if we could get higher number of respondents. Also, majority of the respondents for the Mental health survey belonged to the United States. A new column ('Percent') was computed in this case to avoid incorrect and biased results. Hence, a fair and equal number of respondents from all the countries could have been ideal for such study but also impractical.

7. Bibliography

Bock, T. (2020, January). What is a Correlation Matrix? [Web blog post]. D*isplayr.com*. Retrieved from https://www.displayr.com/what-is-a-correlation-matrix/

Hall, J., & Helliwell, J. F. (2014). Happiness and human development. UNDP Human Development Report: Occasional Paper. New York, USA: United Nations Development Programme.

Helliwell, J., Layard, R., & Sachs, J. (2015). World Happiness Report 2016, Update (Vol. I). New York: Sustainable Development Solutions Network. Retrieved from https://worldhappiness.report/ed/2015/

Helliwell, J., Layard, R., & Sachs, J. (2016). World Happiness Report 2016, Update (Vol. I). New York: Sustainable Development Solutions Network. Retrieved from https://worldhappiness.report/ed/2016/

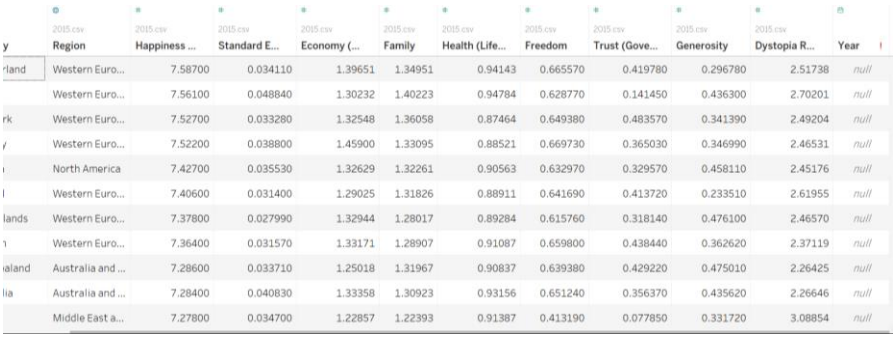
Helliwell, J., Layard, R., & Sachs, J. (2017). World Happiness Report 2016, Update (Vol. I). New York: Sustainable Development Solutions Network. Retrieved from https://worldhappiness.report/ed/2017/

Nwegbu, S. (2015, December). How do I handle a large number of 'not-applicable' in a complex survey dataset? [Web blog post]. *ResearchGate*.net. Retrieved from https://www.researchgate.net/post/How_do_I_handle_a_large_number_of_not-applicable_in_a_complex_survey_dataset_Nigeria_DHS_2008

Osmihelp.org. 2020. Research :: Open Sourcing Mental Illness - Changing How We Talk About Mental Health In The Tech Community - Stronger Than Fear. [online] Retrieved from https://osmihelp.org/research

Appendix

Happiness report statistical dataset for the year 2015

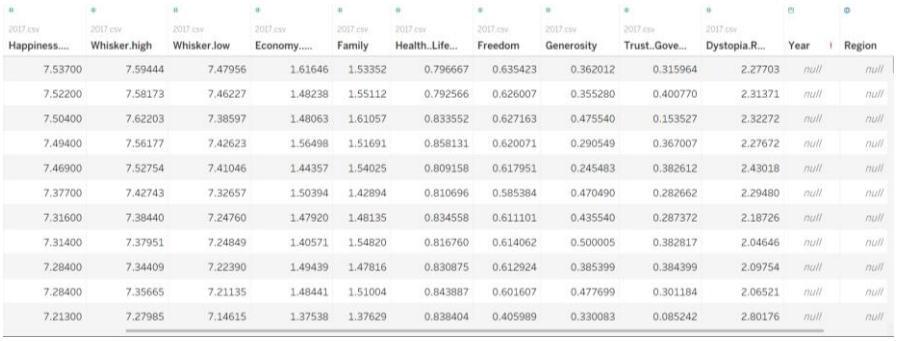


y	Region	Happiness ...	Standard E...	Economy (...	Family	Health (Life...	Freedom	Trust (Gove...	Generosity	Dystopia R...	Year
rland	Western Euro...	7.58700	0.034110	1.39651	1.34951	0.94143	0.665570	0.419780	0.296780	2.51738	null
	Western Euro...	7.56100	0.048840	1.30232	1.40223	0.94784	0.628770	0.141450	0.436300	2.70201	null
rk	Western Euro...	7.52700	0.033280	1.32548	1.36058	0.87464	0.649380	0.483570	0.341390	2.49204	null
y	Western Euro...	7.52200	0.038800	1.45900	1.33095	0.88521	0.669730	0.365030	0.346990	2.46531	null
	North America	7.42700	0.035530	1.32629	1.32261	0.90563	0.632970	0.329570	0.458110	2.45176	null
l	Western Euro...	7.40600	0.031400	1.29025	1.31826	0.88911	0.641690	0.413720	0.233510	2.61955	null
lands	Western Euro...	7.37800	0.027990	1.32944	1.28017	0.89284	0.615760	0.318140	0.476100	2.46570	null
n	Western Euro...	7.36400	0.031570	1.33171	1.28907	0.91087	0.659800	0.438440	0.362620	2.37119	null
aland	Australia and ...	7.28600	0.033710	1.25018	1.31967	0.90837	0.639380	0.429220	0.475010	2.26425	null
lia	Australia and ...	7.28400	0.040830	1.33358	1.30923	0.93156	0.651240	0.356370	0.435620	2.26646	null
	Middle East a...	7.27800	0.034700	1.22857	1.22393	0.91387	0.413190	0.077850	0.331720	3.08854	null

Happiness report statistical dataset for the year 2016

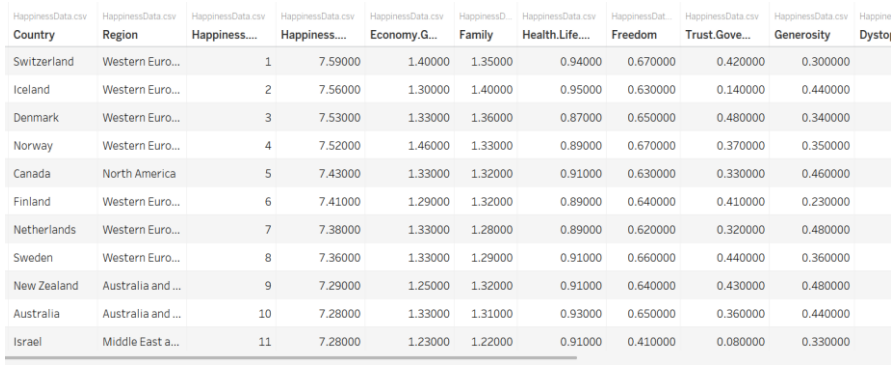
	Happiness ...	Lower Conf...	Upper Conf...	Economy (...	Family	Health (Life...	Freedom	Trust (Gove...	Generosity	Dystopia R...	Year
n Euro...	7.52600	7.46000	7.59200	1.44178	1.16374	0.795040	0.579410	0.444530	0.361710	2.73939	null
n Euro...	7.50900	7.42800	7.59000	1.52733	1.14524	0.863030	0.585570	0.412030	0.280830	2.69463	null
n Euro...	7.50100	7.33300	7.66900	1.42666	1.18326	0.867330	0.566240	0.149750	0.476780	2.83137	null
n Euro...	7.49800	7.42100	7.57500	1.57744	1.12690	0.795790	0.596090	0.357760	0.378950	2.66465	null
n Euro...	7.41300	7.35100	7.47500	1.40598	1.13464	0.810910	0.571040	0.410040	0.254920	2.82596	null
America	7.40400	7.33500	7.47300	1.44015	1.09610	0.827600	0.573700	0.313290	0.448340	2.70485	null
n Euro...	7.33900	7.28400	7.39400	1.46468	1.02912	0.812310	0.552110	0.299270	0.474160	2.70749	null
lia and ...	7.33400	7.26400	7.40400	1.36066	1.17278	0.830960	0.581470	0.419040	0.494010	2.47553	null
lia and ...	7.31300	7.24100	7.38500	1.44443	1.10476	0.851200	0.568370	0.323310	0.474070	2.54650	null
n Euro...	7.29100	7.22700	7.35500	1.45181	1.08764	0.831210	0.582180	0.408670	0.382540	2.54734	null
East a...	7.26700	7.19900	7.33500	1.33766	0.99537	0.849170	0.364320	0.087280	0.322880	3.31029	null

Happiness report statistical dataset for the year 2017



Happiness....	Whisker.high	Whisker.low	Economy.....	Family	Health..Life...	Freedom	Generosity	Trust..Gove...	Dystopia.R...	Year	Region
7.53700	7.59444	7.47956	1.61646	1.53352	0.796667	0.635423	0.362012	0.315964	2.27703	null	null
7.52200	7.58173	7.46227	1.48238	1.55112	0.792566	0.626007	0.355280	0.400770	2.31371	null	null
7.50400	7.62203	7.38597	1.48063	1.61057	0.833552	0.627163	0.475540	0.153527	2.32272	null	null
7.49400	7.56177	7.42623	1.56498	1.51691	0.858131	0.620071	0.290549	0.367007	2.27672	null	null
7.46900	7.52754	7.41046	1.44357	1.54025	0.809158	0.617951	0.245483	0.382612	2.43018	null	null
7.37700	7.42743	7.32657	1.50394	1.42894	0.810696	0.585384	0.470490	0.282662	2.29480	null	null
7.31600	7.38440	7.24760	1.47920	1.48135	0.834558	0.611101	0.435540	0.287372	2.18726	null	null
7.31400	7.37951	7.24849	1.40571	1.54820	0.816760	0.614062	0.500005	0.382817	2.04646	null	null
7.28400	7.34409	7.22390	1.49439	1.47816	0.830875	0.612924	0.385399	0.384399	2.09754	null	null
7.28400	7.35665	7.21135	1.48441	1.51004	0.843887	0.601607	0.477699	0.301184	2.06521	null	null
7.21300	7.27985	7.14615	1.37538	1.37629	0.838404	0.405989	0.330083	0.085242	2.80176	null	null

Combined final Happiness report statistical dataset for the years 2015, 2016 and 2017



Country	Region	Happiness....	Happiness....	Economy.G...	Family	Health.Life....	Freedom	Trust.Gove...	Generosity	Dysto
Switzerland	Western Euro...	1	7.59000	1.40000	1.35000	0.94000	0.670000	0.420000	0.300000	
Iceland	Western Euro...	2	7.56000	1.30000	1.40000	0.95000	0.630000	0.140000	0.440000	
Denmark	Western Euro...	3	7.53000	1.33000	1.36000	0.87000	0.650000	0.480000	0.340000	
Norway	Western Euro...	4	7.52000	1.46000	1.33000	0.89000	0.670000	0.370000	0.350000	
Canada	North America	5	7.43000	1.33000	1.32000	0.91000	0.630000	0.330000	0.460000	
Finland	Western Euro...	6	7.41000	1.29000	1.32000	0.89000	0.640000	0.410000	0.230000	
Netherlands	Western Euro...	7	7.38000	1.33000	1.28000	0.89000	0.620000	0.320000	0.480000	
Sweden	Western Euro...	8	7.36000	1.33000	1.29000	0.91000	0.660000	0.440000	0.360000	
New Zealand	Australia and ...	9	7.29000	1.25000	1.32000	0.91000	0.640000	0.430000	0.480000	
Australia	Australia and ...	10	7.28000	1.33000	1.31000	0.93000	0.650000	0.360000	0.440000	
Israel	Middle East a...	11	7.28000	1.23000	1.22000	0.91000	0.410000	0.080000	0.330000	

Mental Health survey raw dataset

Timestamp	Age	Gender	Country	self_emplo...	family_hist...	treatment	work_interfere	no_employ...	remote_work	tech_co
28/08/2014 10:07:53 AM	-1,726	male	United Kingd...	No	No	Yes	Sometimes	26-100	No	No
27/08/2014 12:39:14 PM	-29	Male	United States	No	No	No	NA	More than 10...	Yes	No
28/08/2014 10:35:55 AM	5	Male	United States	No	No	No	NA	100-500	No	Yes
29/08/2014 9:10:58 AM	8	A little about you	Bahamas, The	Yes	Yes	Yes	Often	1-5	Yes	Yes
27/08/2014 4:16:51 PM	18	male	Finland	No	No	No	NA	500-1000	Yes	No
27/08/2014 2:11:55 PM	18	Male	United States	No	No	No	Never	26-100	No	Yes
27/08/2014 12:15:11 PM	18	something kind...	Russia	No	No	No	NA	26-100	Yes	Yes
27/08/2014 2:10:15 PM	18	male	United States	No	No	Yes	Sometimes	6-25	No	Yes
27/08/2014 11:43:10 AM	18	Male	Netherlands	No	No	No	Often	6-25	No	Yes
27/08/2014 4:19:05 PM	18	Female	United Kingd...	No	Yes	Yes	Sometimes	1-5	Yes	Yes
27/08/2014 12:31:43 PM	18	Male	United States	No	No	Yes	Rarely	1-5	Yes	Yes

Mental Health survey final dataset

Year	Age	Gender	Country	self_emplo...	family_hist...	treatment	work_interfere	remote_work	tech_comp...	benefits	wellness_
2014	31	Male	United States	No	No	Yes	Sometimes	Yes	Yes	Don't know	No
2014	29	Male	Bulgaria	No	No	No	Never	Yes	Yes	Don't know	No
2014	42	Female	United States	No	Yes	Yes	Sometimes	No	No	Yes	No
2014	36	Male	United States	No	Yes	No	Never	No	Yes	Don't know	No
2014	27	Male	Canada	No	No	No	Never	No	Yes	Don't know	Don't k
2014	29	Female	United States	No	Yes	Yes	Rarely	No	Yes	Yes	No
2014	23	Male	United Kingd...	No	No	Yes	Sometimes	Yes	Yes	Don't know	Don't k
2014	32	Male	United States	No	No	Yes	Sometimes	No	Yes	Yes	No
2014	46	Male	United States	Yes	Yes	No	Sometimes	Yes	Yes	Yes	Yes
2014	36	Male	France	Yes	Yes	No	Don't know	Yes	Yes	No	Yes
2014	29	Male	United States	No	Yes	Yes	Sometimes	No	Yes	Yes	No

Final merged Happiness report and mental health survey dataset with computed column 'Percent'

Country (Gr...	Percent	seek_help	Total	Country	Region	Happ...	Happiness....	Economy.G...	Family	Health.Life....	Freedom	Tru
Australia	14.2900	Don't know	21	Australia	Australia and ...	10	7.28000	1.48000	1.51000	0.840000	0.600000	
Australia	14.2900	Don't know	21	Australia	Australia and ...	9	7.31000	1.44000	1.10000	0.850000	0.570000	
Australia	14.2900	Don't know	21	Australia	Australia and ...	10	7.28000	1.33000	1.31000	0.930000	0.650000	
Australia	38.1000	Yes	21	Australia	Australia and ...	10	7.28000	1.48000	1.51000	0.840000	0.600000	
Australia	38.1000	Yes	21	Australia	Australia and ...	9	7.31000	1.44000	1.10000	0.850000	0.570000	
Australia	38.1000	Yes	21	Australia	Australia and ...	10	7.28000	1.33000	1.31000	0.930000	0.650000	
Australia	47.6200	No	21	Australia	Australia and ...	10	7.28000	1.48000	1.51000	0.840000	0.600000	
Australia	47.6200	No	21	Australia	Australia and ...	9	7.31000	1.44000	1.10000	0.850000	0.570000	
Australia	47.6200	No	21	Australia	Australia and ...	10	7.28000	1.33000	1.31000	0.930000	0.650000	
Austria	33.3300	No	3	Austria	Western Euro...	13	7.01000	1.49000	1.46000	0.820000	0.570000	
Austria	33.3300	No	3	Austria	Western Euro...	12	7.12000	1.45000	1.08000	0.810000	0.540000	