**Exploratory Data Analysis**

Basic Information of dataset

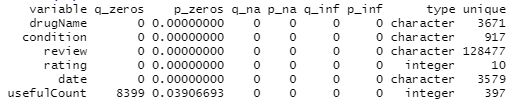
Source -> UCI Machine learning repository

https://archive.ics.uci.edu/ml/datasets/Drug+Review+Dataset+%28Drugs.com%29#

The dataset provides patient reviews on specific drugs along with related conditions and a 10 star patient rating reflecting overall patient satisfaction.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Dataset Info** |  |  | **Variable Types in Dataset** | **Variable Types as Data Science** | **Type** |
| No of variables | 7 |  | chr | drugName | categorical |
| No of observations | 215072 |  | int | condition | categorical |
| Total NA records | 82 |  |  | review | text |
| Total Size in memory | 105.5 MB |  |  | rating | numerical |
|  |  |  |  | date | date |
|  |  |  |  | usefulCount | numerical |
|  |  |  |  | uniqueID | dropped |

Statistics of dataset



Files of Drugs Name and Condition

(1) drugfrequency.csv -> this file gives you count of every drug and its percentage in overall dataset.

(2) drugcondition.csv -> this file gives you count of every condition and its percentage in overall dataset.

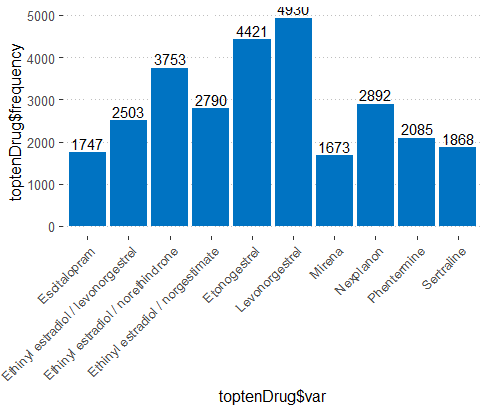
Q1.What is the most popular drug?

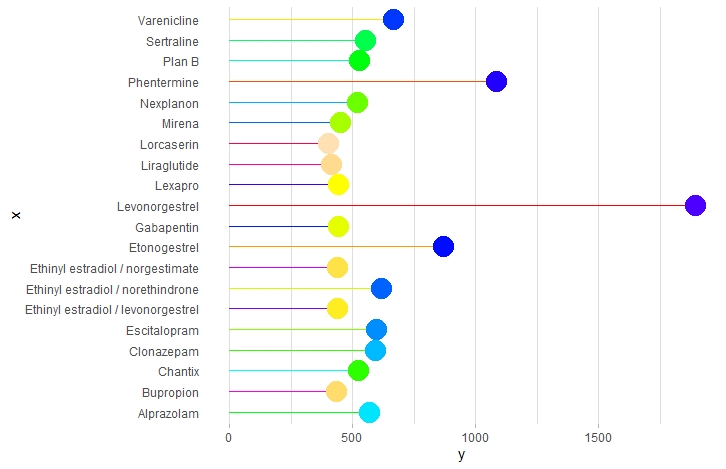
|  |  |  |  |
| --- | --- | --- | --- |
| Levonorgestrel | 4930 | 2.29 | 2.29 |
|  |  |  |  |

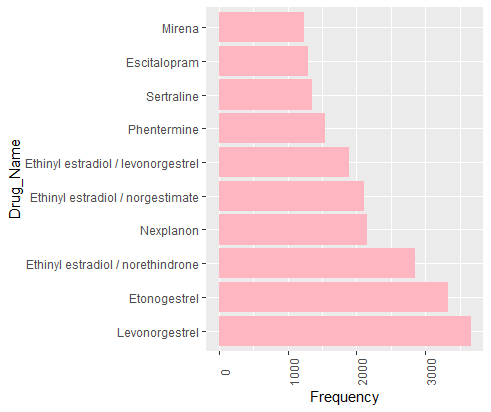
Q2. What is the most popular condition?

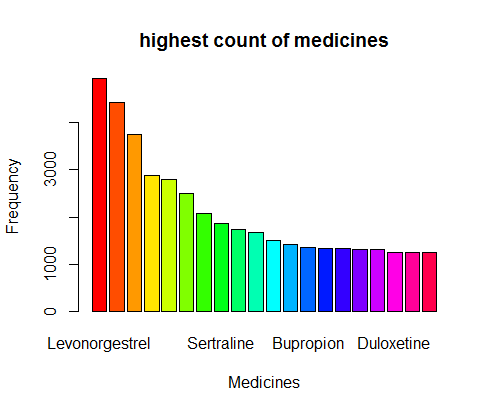
|  |  |  |  |
| --- | --- | --- | --- |
| var | frequency | percentage | cumulative\_perc |
| Birth Control | 38436 | 17.88 | 17.88 |

Top 10 Drugs

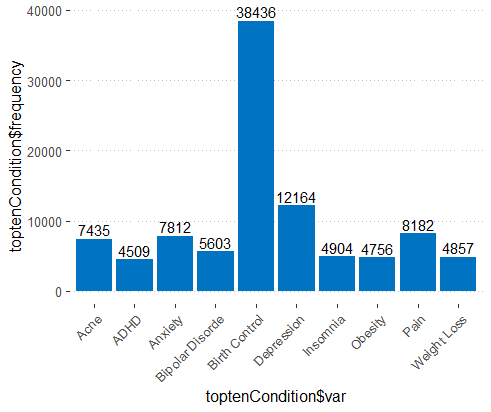


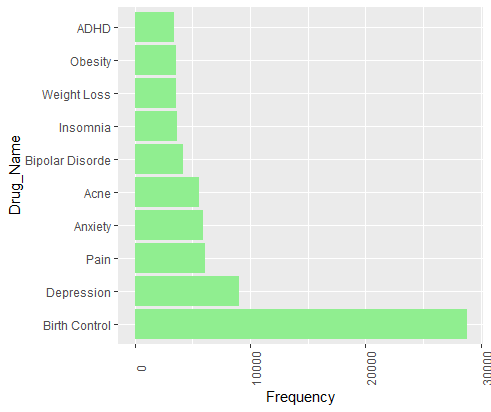




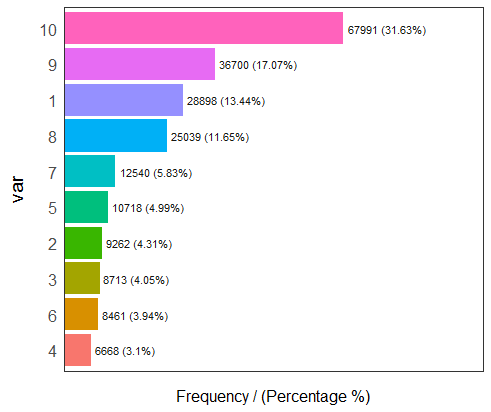


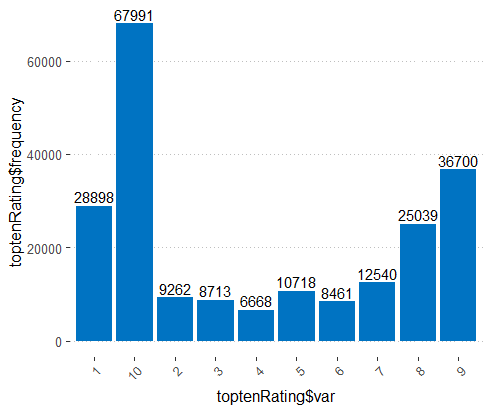
Top Ten Conditions



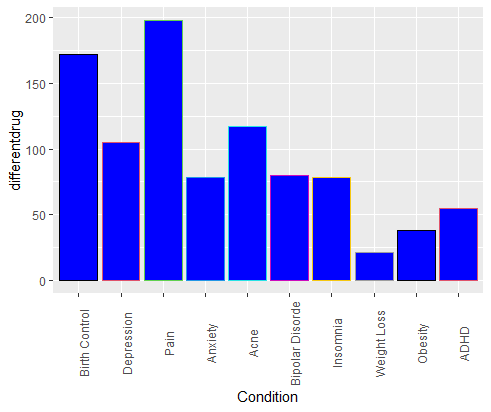


Top Ten Rating

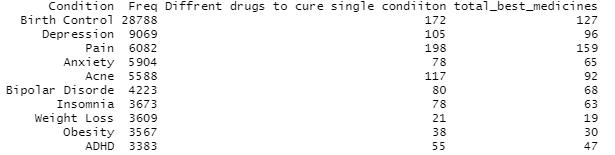




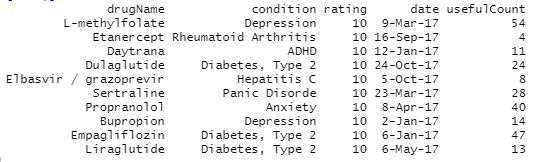
Top Ten condition, individually containing different drugs on the base of condition frequency



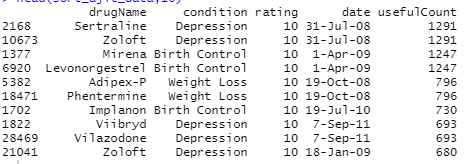
How many top rated (having rating 10) drugs used in Best 10 conditions.



Drugs used in 2017



Top Ten Drugs on the basis of usefulCount

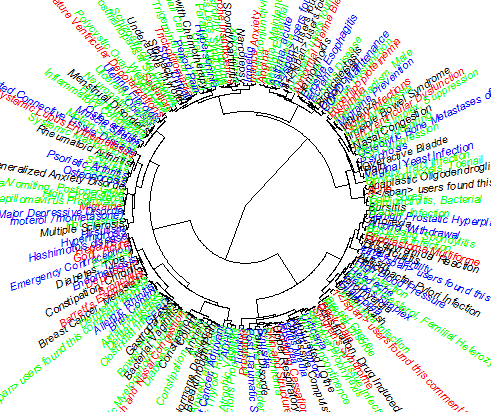


Steps

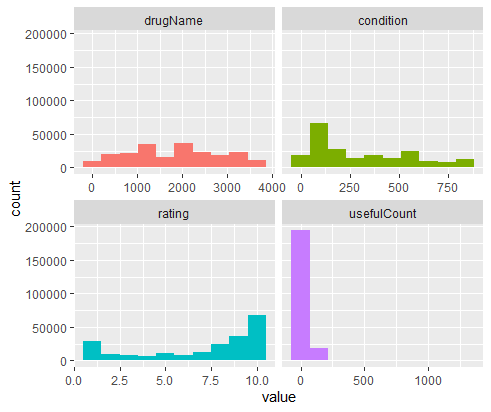
1. Fetch all drugs which has given 10 rating

2. Fetch top ten records from above matrix on the basis of usefulcountes

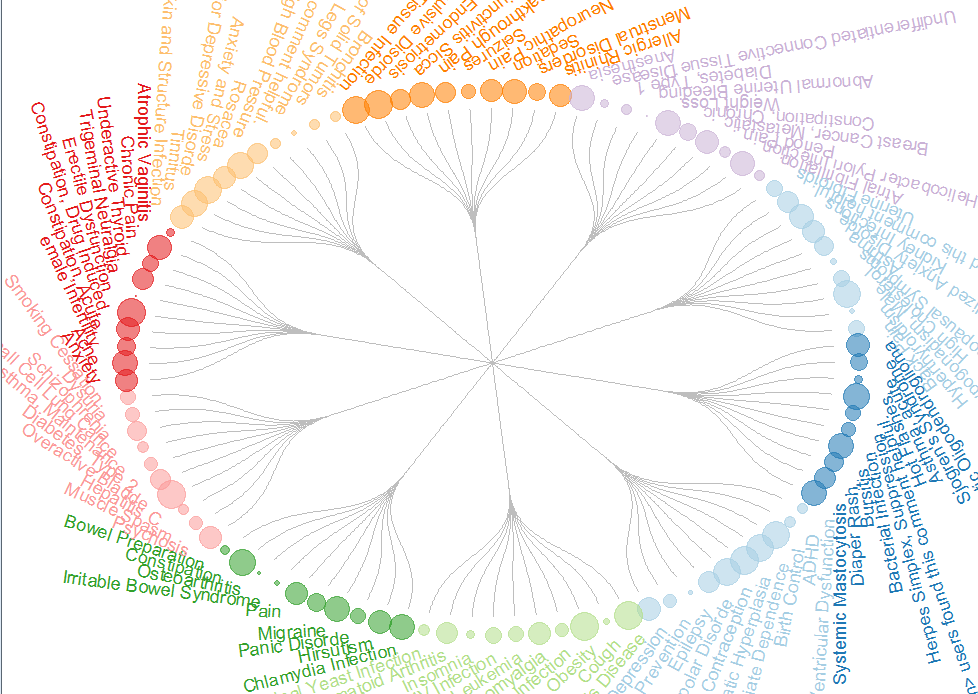
Conditions Plot on Fan style



All Bar Plot in single frame



Customize Circle Dendrogram



Remaining Items to do as a part of EDA

(1) Eliminating false data in dataset

Done

