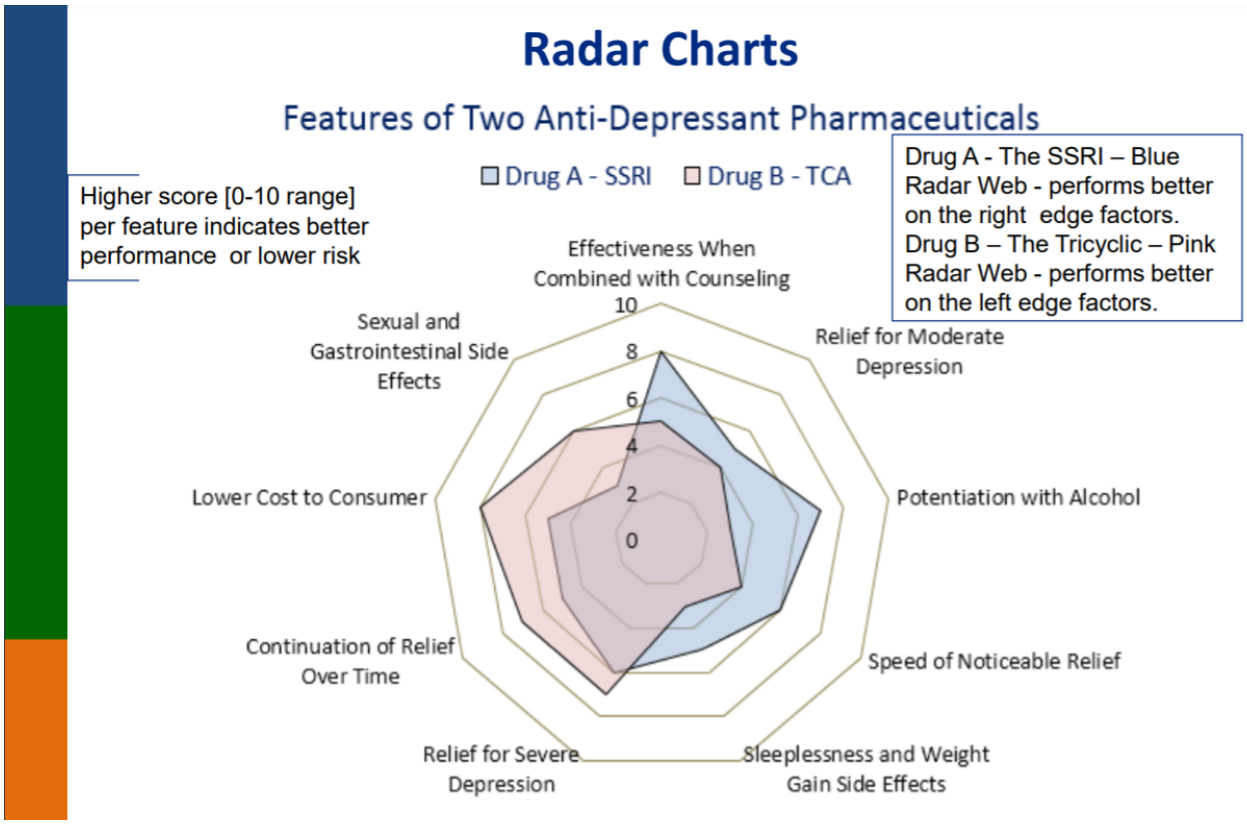
1. Visit the website https://datavizproject.com/ and pick one data visualization/plot that interests you. Find out how it is drawn and what aspects of the data the different components represent. Be prepared to explain how your visualization of choice works in the class.

Radar Diagram

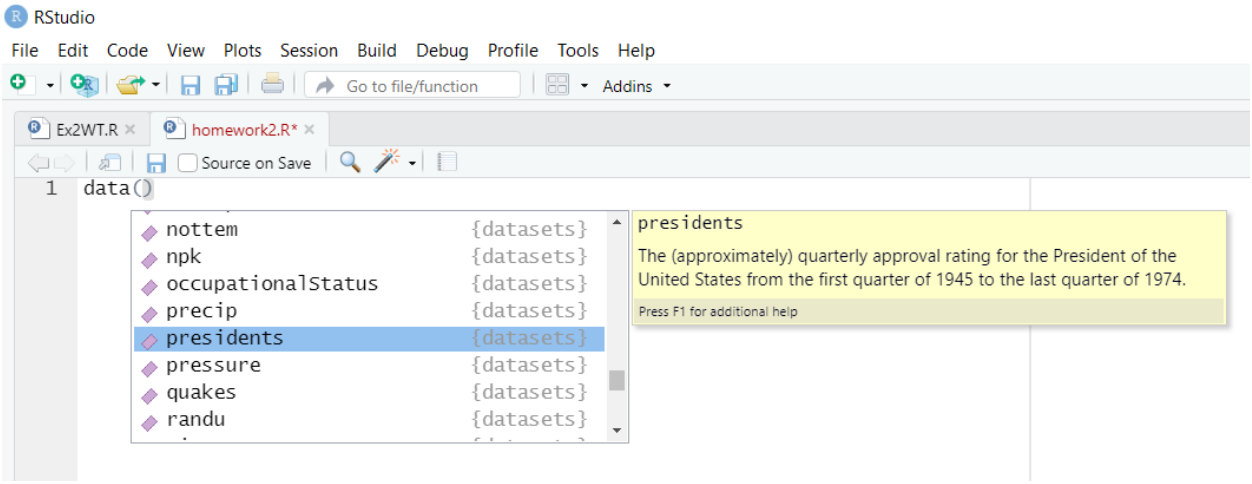


Generally do not attempt to compare more than three groups on one radar or web chart.
And do not attempt to display more than ten factors on one radar or web chart.

- Radar Charts are used to compare two or more items or groups on various features or characteristics. Example: Compare two anti-depressant drugs on features such as efficacy for severe depression, prevalence of specific side effects, interaction with alcohol, continuation of relief over time, cost to the consumer etc.
- Typically the features or factors to be compared are rather different or disparate from each other.
- Often the scores assigned to each factor are relatively scaled – for example, 0-10, where higher scores indicate better performance or lower risk on the factor under consideration.
- Scores on each factor or feature radiate outward on spokes from a central zero hub.
- The scores for the factors for each group (anti-depressants, in this example) are connected to form a “radar image” or “spider web” pattern.
- The patterns for each group often overlap, so transparent shading of the group patterns and sorting of the scores by group will aid in the visual display of the group “webs” or “radar sweeps”.

2. Type the command data() in R to show all data sets currently available in your installed packages. Go through the data sets and pick one that interests you. Check the help file of the data set using the command ?packagename for more detailed information. Be prepared to describe your answers to the following questions in the class:

What is the purpose of the data? What kind of phenomenon does it describe?



The (approximately) quarterly approval rating for the American President from first quarter of 1945 to last quarter of 1974

What kind of study is behind the data (observational, controlled, simulation, survey or something else)?

It is an observational study that collects data about approval ratings. However, this dataset is a fudged or somewhat misinterpreted version of the approval ratings. Source is from The Gallup Organization.

How is the data represented in R (univariate, multivariate, time series…)?

This data package is a time series of 120 values.

What kind of plots would you use to best summarize the data?

Since this is a time series, it is best to represent the data with Line Graph. It is intuitive and helps us get a quick sense of how US approval rating has changed over 30 years

What kind of numerical statistics would you use to best summarize the data?

The mean, median, mode, percentiles, range, variance, and standard deviation are the most commonly used numerical measures for quantitative data.