Time Series

Xiaotai, David, Lydia, Michael

Get the data

Data at -

https://www.transtats.bts.gov/DL_SelectFields.asp?Table_ID=236&DB_Short_Name=On-Time

https://cdn.rawgit.com/mikejt33/DataViz/246c2026/data/flights.csv.gz

flights.csv

"DAY_OF_WEEK": 1-7
"FL_DATE": 2017-MM-DD
"UNIQUE_CARRIER": XX
"AIRLINE_ID": 12345
"CARRIER": XX
"TAIL_NUM": N123XX
"FL_NUM": 1234
"ORIGIN": ABC
"DEST": ABC
"DEP_TIME": 1234
"DEP_DELAY": negative or positive integer
"DEP_DEL15": 0, 1

"ARR_DELAY: negative or positive integer
"ARR_DEL15": 0, 1
"CANCELLED": 0, 1
"CANCELLATION_CODE": NA or code
"DIVERTED": 0, 1
"AIR_TIME": some integer
"DISTANCE": some integer

Time Series Terms

Decomposition - task that deconstructs a time series into several components, each representing one of the underlying categories of patterns.

Additive vs Multiplicative https://anomaly.io/seasonal-trend-decomposition-in-r/

Detrending - is removing a trend from a **time series**; a trend usually refers to a change in the mean over **time**

Time Series Terms

Smoothing

Moving Averages: For each value in the discrete time index, *t*, compute the average (simple or weighted) of the observed responses within a specified neighborhood centered at the current index.

Local Regression (LOESS): For each value in the discrete time index, *t*, compute the fit using a weighted subset of the data within a specified neighborhood of the current index and use the model to compute the response.

Why Smoothing

Smoothing is usually done to help us better see patterns, trends for example, in time series. Generally smooth out the irregular roughness to see a clearer signal. For seasonal data, we might smooth out the seasonality so that we can identify the trend. Smoothing doesn't provide us with a model, but it can be a good first step in describing various components of the series.

Creating Time Series Objects in R

- We can create Time Series Objects in R using the ts function

The function ts is used to create time-series objects.

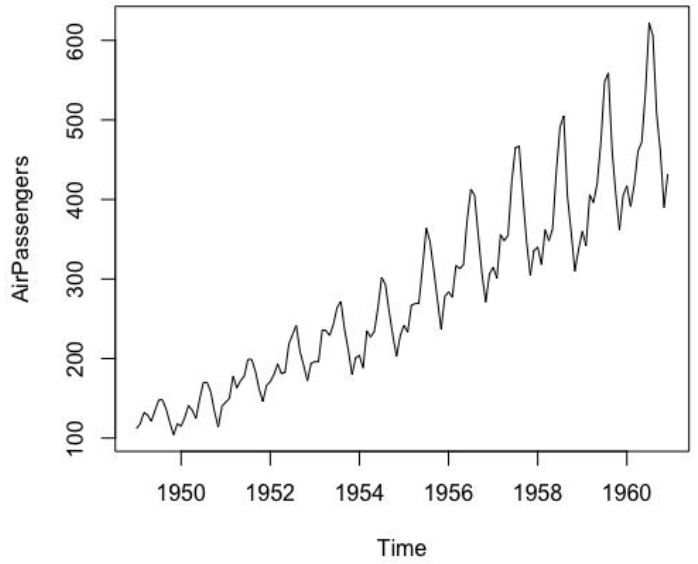
as.ts and is.ts coerce an object to a time-series and test whether an object is a time series.

https://www.rdocumentation.org/packages/stats/versions/3.4.3/topics/ts

TS Object Example

```
> AirPassengers
     Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
1949 112 118 132 129 121 135 148 148 136 119 104 118
1950 115 126 141 135 125 149 170 170 158 133 114 140
1951 145 150 178 163 172 178 199 199 184 162 146 166
1952 171 180 193 181 183 218 230 242 209 191 172 194
1953 196 196 236 235 229 243 264 272 237 211 180 201
1954 204 188 235 227 234 264 302 293 259 229 203 229
1955 242 233 267 269 270 315 364 347 312 274 237 278
1956 284 277 317 313 318 374 413 405 355 306 271 306
1957 315 301 356 348 355 422 465 467 404 347 305 336
1958 340 318 362 348 363 435 491 505 404 359 310 337
1959 360 342 406 396 420 472 548 559 463 407 362 405
1960 417 391 419 461 472 535 622 606 508 461 390 432
```

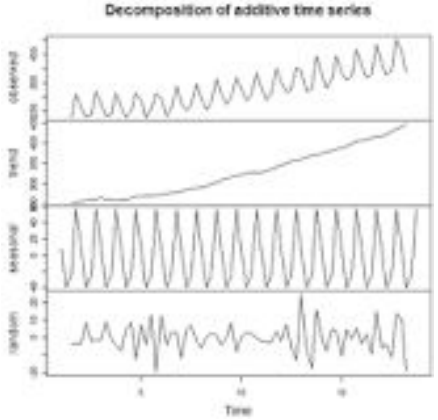
Time Series Plot



Decomposition

The **decomposition** of **time series** is a statistical task that deconstructs a **time series** into several components, each representing one of the underlying categories of patterns.

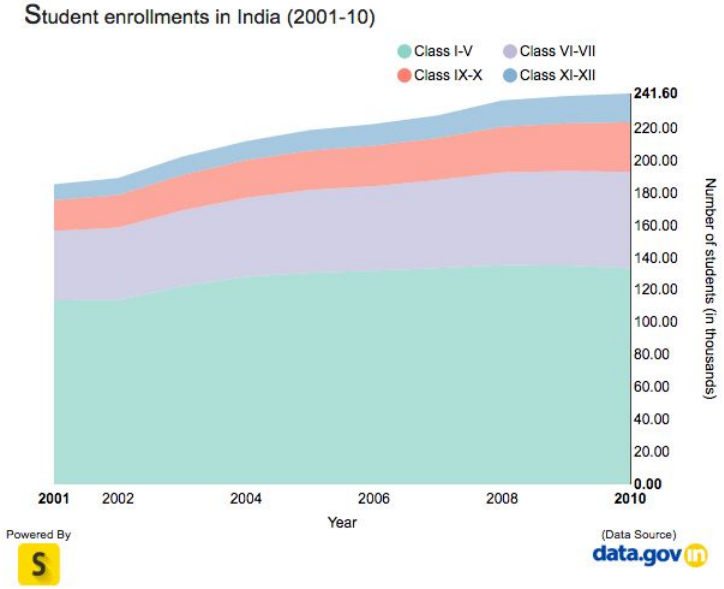
Example Output:



Stacked Area Chart

An area chart is similar to a line chart in that it has points connected by straight lines on a two-dimensional chart. It also puts time as the independent variable on the x-axis and the dependent variable on the y-axis. However, in an area chart, multiple variables are “stacked” on top of each other, and the area below each line is colored to represent each variable. Stacked area charts are useful to show how both a cumulative total and individual components of that total changed over time.

-This is a stacked area chart showing time series data of student enrollments in India from 2001-10.

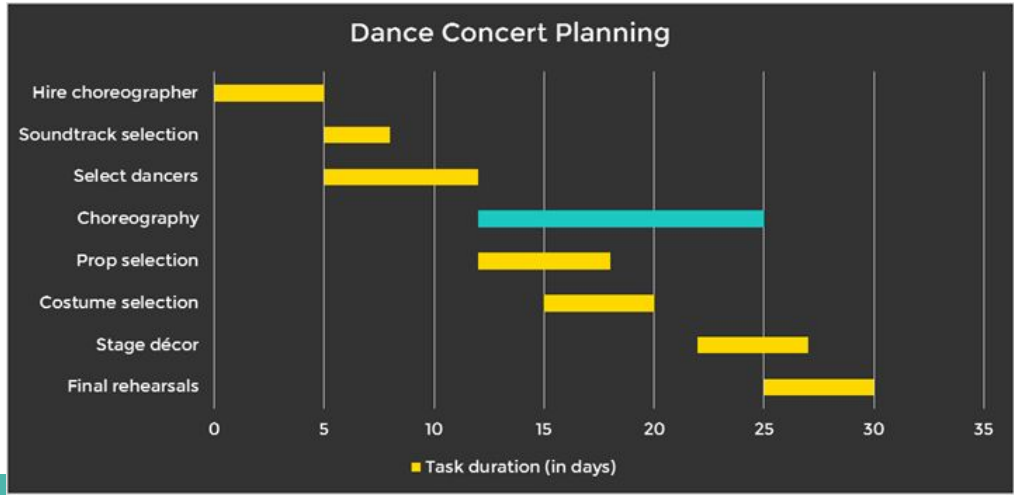


- R-code step by step for creating a Stacked Area Chart https://www.r-graph-gallery.com/136-stacked-area-chart/

Gantt Chart

- A Gantt chart is a horizontal bar chart showing work completed in a certain period of time with respect to the time allocated for that particular task.
- Gantt charts can help you plan for complex, long-term projects that are likely to undergo several revisions and have various resource and task dependencies.
- Useful when tracking lots of activities to be completed, some of which will take place simultaneously while some can be done only after another activity has been completed.
- You can show additional information such as the correlation between individual tasks, resources used in each task, overlapping resources, etc.

-Here's a good tutorial depicting how to create Gantt Charts in R
https://davetang.org/muse/2017/02/03/gantt-chart-using-r/

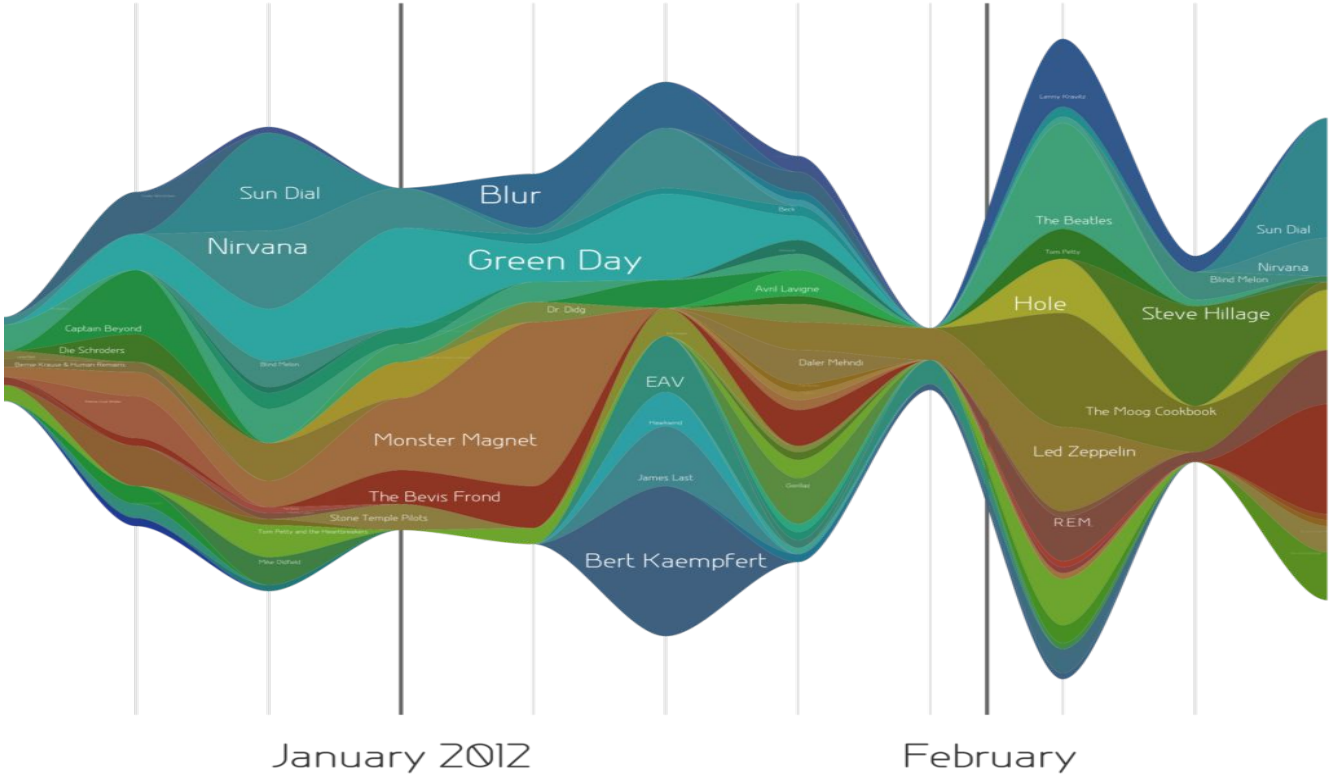


Stream Graph

- A stream graph is essentially a stacked area graph, but displaced around a central horizontal axis. The stream graph looks like flowing liquid, hence the name.
- Stream graphs are great to represent and compare time series data for multiple variables. Stream graphs are, thus, apt for large data sets.
- Variables that do not have significantly high values might tend to get drowned out in the visualization if the colors are not chosen well.

- Here's a good tutorial depicting how to create Stream Graphs in R
  - https://www.r-bloggers.com/introducing-the-streamgraph-htmlwidget-r-package/

Example of Stream Graph



Heat Maps

- Geospatial visualizations often use heat maps since they quickly help identify “hot spots” or regions of high concentrations of a given variable.
    - When adapted to temporal visualizations, heat maps can help us explore two levels of time in a 2D array.
- Heat maps are perfect for a two-tiered time frame — for instance, 7 days of the week spread across 52 weeks in the year, or 24 hours in a day spread across 30 days of the month, and so on.
- The limitation, though, is that only one variable can be visualized in a heat map. Comparison between two or more variables is very difficult to represent.

- R - Code tutorial for creating a Timeseries Heat-map:
    - https://www.r-bloggers.com/ggplot2-time-series-heatmaps/

HeatMap Example

This heat map visualizes birthdays for babies born in the United States between 1973 and 1999. The vertical axis represents the 31 days in a month while the horizontal axis represents the 12 months in a year. This chart quickly helps us identify that a large number of babies were born in the later half of July, August, and September.