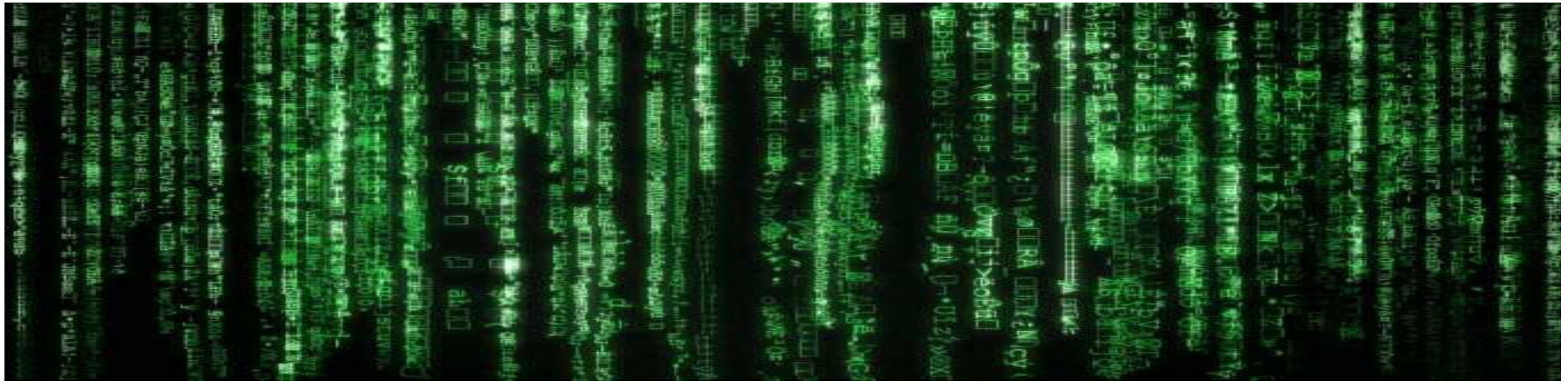


INF 110 Discovering Informatics

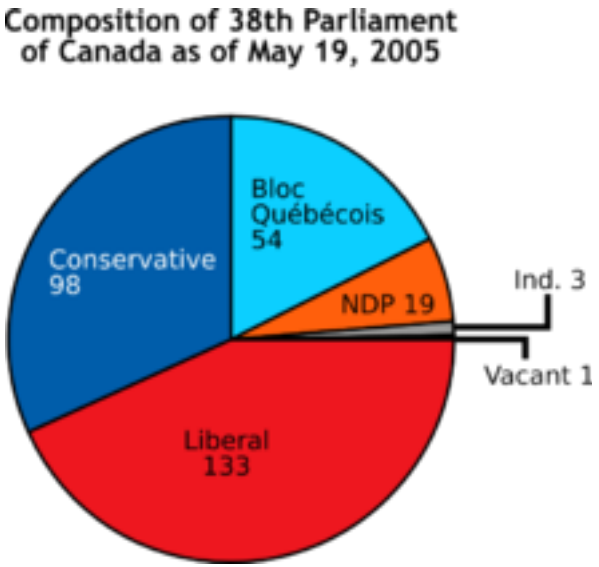
Charts

And when (but not necessarily how) to use them

NAU
NORTHERN
ARIZONA
UNIVERSITY

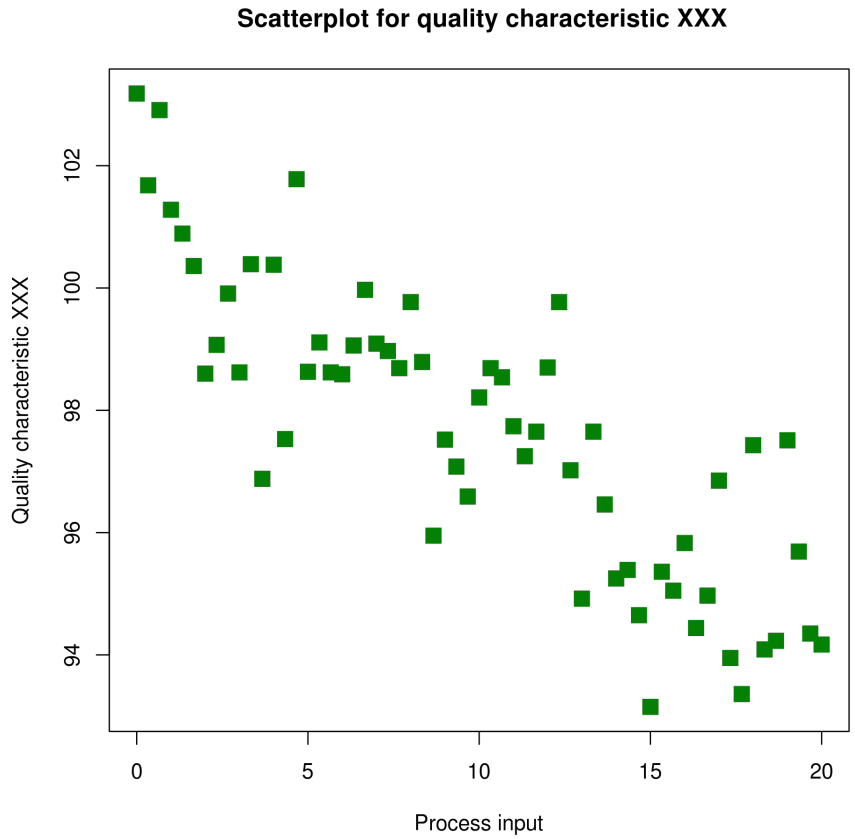
Charts

- A graphical representation of data
- Humans are not good at seeing trends and drawing conclusions from data.
- Often easier to compare graphics than raw numbers
- Sometimes called "graphs" or "plots"



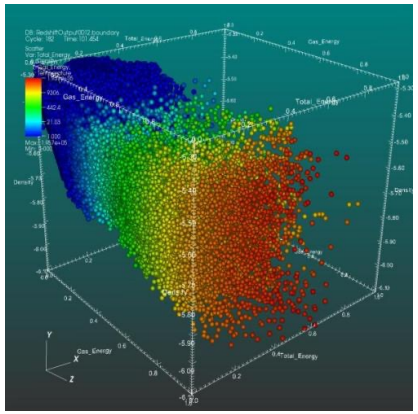
Scatter Chart

- Two dimensions, or variables
- Places a dot at the intersection of two variables
- Uses x-y coordinates to display information
- Each axis represents a variable
- Can only support two variables



Hacks

- 3D plots can support three variables
- Different colors and symbols can support more variables (but harder to read)



Boiler Plate Code (Part 1)

```
# Without this line you won't see your plots in Jupyter:
%matplotlib inline
```

Boiler Plate Code (Part 2)

```
# Make your plots look fancier:

import matplotlib.pyplot as plots
plots.style.use('fivethirtyeight')
```

https://fivethirtyeight.com/

Scatter Plots in Python

- We will use the scatter method in datascience
- You could use the scatter function for matplotlib, too (but stick with datascience).

- Call the table.scatter method
- Supply two variables (the x and y labels)
- The data in each column is then grabbed by the functions it does the work

```
t.scatter('1st variable', '2nd variable')
```

```
Note that t is a datascience Table object!
```

Live Code Actors

Tasks: Using the actors data set, create a scatter plot of:

- Number of Movies
- Total Gross

then remove outliers from the plot.

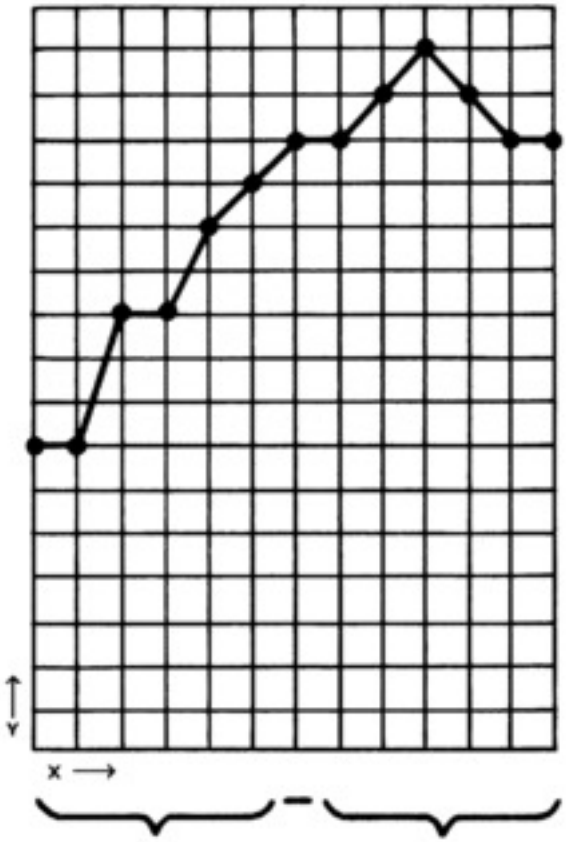
Learning Outcomes

- Creating scatter plots
- Interpreting scatter plots

Line Chart

Same idea as scatter plot except:

- One axis is ordered (usually x)
- And connected by lines to indicate an ordered relationship
- Scatter plots are used to see clustering in the data, but line charts are used to see trends
    - Like increases or decreases as a function of time.



Hacks

- Different colors and symbols can support more variables

Line Charts in Python

- Use the plot function.
- Built into the table object
- Provide the two labels (variables) which represent columns in the data.

```
t.plot('1st variable', '2nd variable')
```

Live Code Movies by Year

Tasks: Using the movies by year data set, create a line chart of:

- Year
- Number of Movies

Learning Outcomes

- Creating line plots
- Interpreting line plots

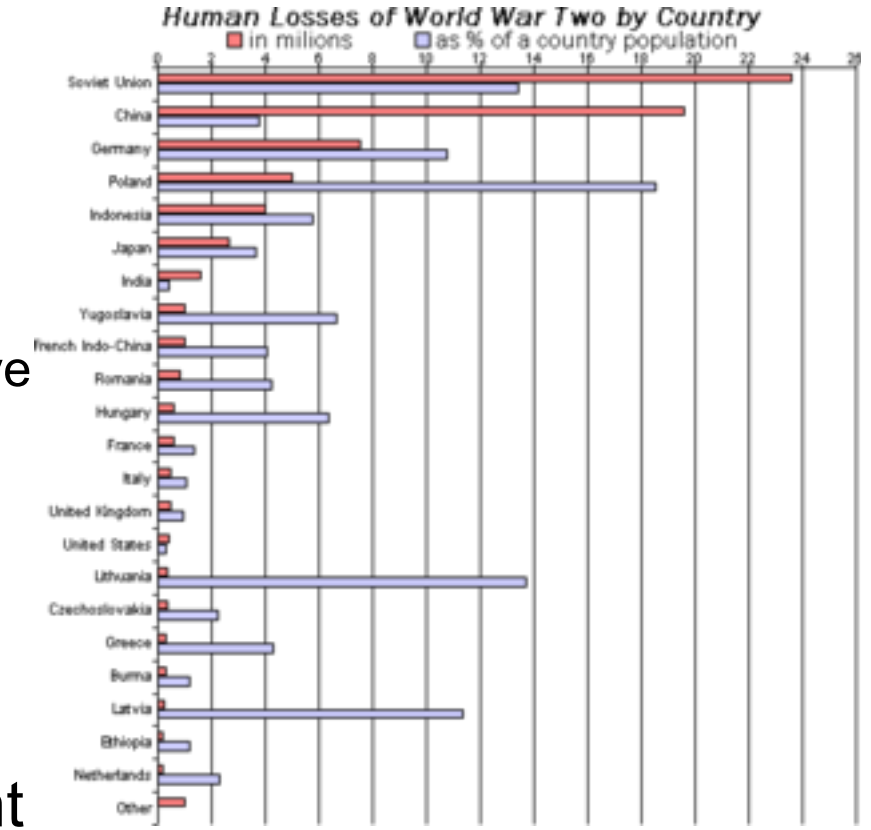
Bar Charts

- Bar charts are a little different.
- In scatter and line charts, both axes are continuous or ordinal.
- In a bar chart, one axis is ***categorical***.

- For instance, if you were looking at cars, the categorical units might be red, blue, and white.

Bar Charts

- Represents categorical data with rectangular bars
- First axis represents categories
- Second axis represents length proportional to value
  - If you wanted to show 10 red cars, the bar would have a unit length of 10.
- Make sure the axis origins are set to zero!
  - That would misrepresent the data.



Hacks

- Bars can be clustered into groups to represent hierarchical data or different conditions

Bar Charts in Python

- Another table based function!
- Use “barh” for horizontal, just “bar” for vertical.

- First argument is the label for your categories of interest
- Second argument is the column label where the frequencies or numbers are.

```
t.barh(‘categories’, 'frequencies')
```

Live Code Class Survey

Tasks: Create a table of the number of times students in last month:

- Streamed a Movie (e.g., Netflix, YouTube)
- Watched a Broadcast Movie (e.g., Cable, OTA)
- Watched on Physical Media (e.g., DVD, Bluray)
- Went to the movie theater

then summarize this data in a bar chart

Learning Outcomes

- Summarizing data
- Creating bar charts
- Interpreting bar charts

Live Code Top Movies

Tasks: Use the group method to summarize receipts for the top 100 movies then summarize this data in a bar chart

Learning Outcomes

- Summarizing data
- Creating bar charts
- Interpreting bar charts


end