Communication and Reporting

DATA 202 21FA

Q&A

> Are regex's useful in any real life dataset?

Q&A

> Are regex's useful in any real life dataset?

	Team
1	Air Force (MWC)
2	Akron (MAC)
3	Alabama (SEC)
4	Appalachian St. (Sun Belt)
5	Arizona (Pac-12)
6	Arizona St. (Pac-12)
7	Arkansas (SEC)
8	Arkansas St. (Sun Belt)
9	Army West Point (FBS Independent)
10	Auburn (SEC)

> "We also tried using a space as the separate factor, but ran into problems with a team like Air Force (MWC) as it would separate the "Air" and "Force" and omit the conference."

	Team
1	Air Force (MWC)
2	Akron (MAC)
3	Alabama (SEC)
4	Appalachian St. (Sun Belt)
5	Arizona (Pac-12)
6	Arizona St. (Pac-12)
7	Arkansas (SEC)
8	Arkansas St. (Sun Belt)
9	Army West Point (FBS Independent)
10	Auburn (SEC)

- Team name is anything except open paren, at least once. `"([^(]+)"`
- Then space and open paren. `" \("`
- Then conference is anything except close paren, at least once. `"([^)]+)"`
- Then close paren. `"\)"`

Q&A

> What's the valid SSN regex?

```
^     # beginning of string
\d{3} # 3 digits
-     # hyphen
\d{2} # 2 digits
-     # hyphen
\d{4} # 4 digits
$     # end of string
```

> Are regexes used in every programming language?

Pretty much. JavaScript even has regex *literals* (`string.match(/regex/)`)

Objectives

- Describe components of an effective data science report
- Contrast a report with a dashboard
- Choose a reporting and presentation strategy appropriate for a given audience

Final Project Expectations

https://cs.calvin.edu/courses/data/202/21fa/projects.html#Project_2_De

Making a Data-Driven Argument

Key points

- Consider the *audience* to get the level of detail right.
  - Never assume your audience can rapidly process complex visuals. (Claus Wilke)
- Consider the *purpose* to choose report vs dashboard vs presentation
- Anchor *claims* in *data*.
- Tell stories (e.g., "but-therefore")

Make a point

Report A

MAIN POINT

- Supporting chart 1
- Supporting table 2
- Supporting model 3

Discussion about how each supports main point

Report B

- Chart 1
- Table 2
- Model 3
- Chart 4
- Table 5
- Chart 6
- Model 7
- Chart 12
- Table 25

Tell a Story

- Chart 1
- Therefore, chart 2
- BUT, chart 3

but-therefore

See also: "Telling a story and making a point"

Anchor conclusions in data

- The units are probably seconds
- The fit looks good
- This was surprising

Anchor conclusions in data

- The units are probably seconds
- The fit looks good
- This was surprising
- because the median, 600, would be 10 minutes
- because the mean error of $15 is less than 0.1% of the price
- because I expected that people would leave higher ratings on products they enjoyed more

Use appropriate language

Plain language for the overview, conclusion, and visuals.

- Labels in visuals: use real names, not `code_names`. (For all aesthetics, not just x and y.)
- Don't assume the reader knows the structure of the data.

Technical language when describing methods (data acquisition, wrangling, modeling, etc.).

- What data representation choices did you make? *why*?
- What modeling choices? Why? etc.

Some color tips

https://blog.datawrapper.de/beautifulcolors/

Start Simple!

Tools for Communication

- Markdown: **know** your *formatting*, including `- lists` and `[links](URL)`.
- Data graphics: `ggplot`, `plotly`, `rbokeh`
- Slides:
  - These slides are `xaringan` + `xaringanthemer` + `xaringanExtra`
  - Other options include `ioslides`, `slidy`, ...
- Getting on the web
  - GitHub Pages
  - `flexdashboard`
  - Shiny apps
  - RStudio Connect

Dashboard vs Report

Source: https://commons.wikimedia.org/wiki/File:Toyota_Corolla_2016_speedometer.jpg

Dashboard vs Report

aspect	Dashboard	Report
shows data?		
uses visuals?		
explanation?		
single page?		

Inspired by https://chartio.com/blog/dashboards-vs-reports-how-theyre-the-same-how-theyre-different/

flexdashboard

https://pkgs.rstudio.com/flexdashboard/articles/examples.html

Example: Shiny Apps

https://shiny.rstudio.com/gallery/

- Engineering Production-Grade Shiny Apps

Analyzing a dashboard

Questions:

1. What are the first things that draw your eye?
2. What information seems to be the most important?
3. For what purpose was this dashboard probably designed?

Examples:

- weather app
- Kent County COVID