Stat7350: Assignment One

Due: March 21, 2019

Invasive Species in Africa dataset

We will explore a dataset on Invasive Species in Africa. This was used as a Tidy Tuesday dataset last year.
The data and information is available here: https://github.com/rfordatascience/tidytuesday/tree/master/data/2018/2018-09-25
The data is originally from a paper published here: https://www-pnas-org.uml.idm.oclc.org/content/113/27/7575

There are multiple tables available:
* `africa_species.csv` contains the known invasive species for African countries from `http://griis.org`
* Table 1: Ranking of all threatened countries by overall invasion threat (`invasion_threat`).
* Table 2: Ranking of all threatened countries by total invasion cost (`invasion_cost`).
* Table 3: Ranking of all threatened countries by total invasion cost (`invasion_cost`) as a proportion (gdp_proportion) of mean GDP (gdp_mean).
* Table 4: Ranking of all source countries by total invasion cost (`invasion_cost`)
* Table 6: List of 140 species and their maximum recorded percentage impact on one of their known host crops (source: CABI Crop Protection Compendium)

All graphics should be created using `ggplot2`. The remainder of your coding is not required to be in the tidyverse style, but this is strongly encouraged.

Start a new `Assignment1` folder in your `.Rproj` directory and push to gihub

The subfolder should be organized in a logical fashion. At minimum it should have a `figures` folder and `data` folder. Download the data from github to the data folder.

Your entire workflow for this assignemnt should be documented in an R Markdown file.

1. Identify a question

Based on the information you have about this dataset, you should clearly identify a broad, interesting question and explain how you arrived at that question.

2. Interact with the data

Make preliminary figures to explore the dataset. For each preliminary figure, explain what pattern the graph is intended to display. For each, explain the flaws in the graph – why does it not succeed? If necessary, update your question of interest and repeat this workflow.

3. Redraw the most useful graph using principles of effective display.

Choose a graph that well captures the story you are trying to tell about this dataset. Document your workflow to design this graph according to principles of good graph design. Explain how your choices effectively display patterns in the data.

4. Save the final graph

Clearly identify a final graph and save it as a pdf in the `figures` folder.

Return to your original question and hypothesis. What have you learned about the data based on this visualization exercise?

5. Identify follow-up questions

Identify additional questions about this dataset that emerge from this exercise. What additional data or information would you require to answer those questions?

Evaluation

Components of the assignment will be graded on a 3-point scale. Note that the *Description of workflow* and *Presentation: final graphic* topics will be double weighted, for the final score (out of 24). The general rubric is here:

Weight	Topic	Excellent: 3	Satisfactory: 2	Needs Work: 1
1	Explanation and justification of question(s)	Questions are testable given the data. The student has described the rationale behind the question, providing context for how they came up with the question.	There is a single focal question that is testable. The student has described the rationale behind the question.	There is a single focal quesiton that relates to the data. The rationale for the question is unclear.
2	Description of workflow	The student provides a thorough description of the workflow used to answer their question. What steps did they take to answer their question. This can include everything from data tidying to visualization to analysis. Justification for the workflow is included in this description.	The student provides an adequate description of the workflow of the analyses used to answer their question.	The student provides a broad-based description of the workflow of their process, but is not able to break it down into specific steps. The description reads as an abstract rather than a concrete set of actions.
2	Presentation: final graphic	Data visualisation is clearly relevant to the question(s) being asked. Figure can be interpreted with minimal additional context (i.e., can stand alone). Axes are well labeled, legends are clear, color schemes make key points easily understandable to the reader. Minutaue of font-sizes, visual aesthetics show clear attention to detail.	Data visualisation is clearly relevant to the question(s) being asked. The visualisation conveys information that is related to the question. There are some problems with aesthetic choices.	Visualization conveys information related to questions but is difficult to interpret.
1	Coding and documentation style	Student has gone beyond what was expected and required, markdown file is well commented and logical.	Markdown style lacks refinement and has some errors, but code is readable and has mostly logical flow.	Many errors in markdown style, little attention paid to making the code human readable.
1	Coding strategy	Complicated problem broken down into sub-problems that are individually much simpler. Code is efficient, correct, and minimal. Code uses appropriate data structure. Code checks for common errors.	Code is correct, but could be edited down to leaner code. Some hacking instead of using suitable data structure. Some checks for errors.	Student does not display the expected level of mastery of the tools and techniques in this course. Chosen task was too limited in scope.
1	Ease of access for instructor, compliance with course conventions for submitted work	Access as easy as possible, code runs!	Satisfactory.	Not an earnest effort to reduce friction and comply with conventions and/or code does not run.