Project #2 – Exoplanet Database Investigation

Proposal Due 3/28 as part of Prelab (Instructions in prelab)
Report, Notebook and Slide Due 4/12
Presentations in class on 4/12 and 4/15

Purpose: Utilize your developing statistical and data visualization skills to design and complete a data-driven investigation of exoplanet properties. You will design the "question" that is the focus of your investigation yourself, in consultation with Prof. Follette and your TAs.

A good question should be:

1) Tractable (the data are capable of answering it)
2) Specific
3) Substantive (the answer is not intuitively obvious)

Constraints/Requirements:

1. You must generate **at least** two well-designed graphics that are informative in the context of your investigation.
2. Your investigation must involve at least three columns of data from the table, and must summarize these quantities quantitatively with summary statistics and/or plots.
3. Your investigation must involve either a fit to a linear trend (see Feb 13 lecture notes), a quantitative evaluation of a hypothesis (see March 6 in class activity and Lab 6), or both.

Products:

1. **A 3-5 page (no more than 5 pages!) written report that includes:**
    a. Background Section
        i. A description of the exoplanet database as a whole, as well as any subset(s) of data that you selected for investigation. Consider inserting plots that you've generated in your labs and prelabs to set the context. Define any new terms (e.g. "radial velocity") in language that a student who has not taken this class and does not know astronomy jargon would understand.
        ii. A description of how you came up with your investigation, and any background information necessary to understand your results.
    b. Procedure Section
        i. A detailed description of any selections or manipulations to the "raw" data from the database that were necessary to complete your investigation
        ii. A description of what you did and why you did it at each step in your investigation/manipulation of the data (for example, why you chose to compute a median over a mean, why you decided to on a particular significance level, etc.).
        iii. In cases where you went beyond what we've done in class in terms of either statistical analyses or computational techniques, clearly explain what you did and why it was necessary to your investigation.

c. Discussion and Analysis Section
    i. Your visualizations, computed statistics, and what they mean, arranged in a narrative with data-driven interpretation of their meanings
    ii. If you've fit a trend: How did you first notice the trend? Did you have to do any manipulations to fit the line (e.g. remove outliers)? What is the physical meaning of the fit? What does it tell you about exoplanets and why does/doesn't it make sense?
    iii. If you've done a hypothesis test: What was your null vs. alternative hypothesis? How did you select a significance level? What does the result tell you about exoplanets and why does/doesn't it make sense?

2. **A jupyter notebook** containing the code needed to read in the data, manipulate it, and create your plots. The notebook must:
    a. Be zipped together with any supplemental files (e.g. data) that are necessary
    b. Execute linearly and without errors
    c. Include a comment describing what each line of code does
3. **A 1-3 slide presentation** highlighting the results of your investigation. It should:
    a. Be submitted as a .pdf
    b. Include minimal text and no complete sentences

A rubric for evaluation of the project is below. ***Please carefully study the criteria against which you will be evaluated before beginning the project.***

	4 points	**7 points**	**10 points**
Background/ Motivation	Student explanation of background had many and/or severe deficiencies in completeness or clarity	Student explanation of background fell a little short in completeness or clarity	Student clearly and thoroughly explained the background necessary to understand their investigation
Procedure	There were many and/or severe deficiencies in the clarity or thoroughness of the procedure explanation or the nature of that procedure	There were some deficiencies in the clarity or thoroughness of the procedure explanation or the nature of that procedure	The procedure used to analyze the data and generate the plots was clearly and thoroughly explained and justified
Final Graphics	Graphics had several and/or severe deficiencies in design or appropriateness to the investigation	Graphics had minor deficiencies in design or appropriateness to the investigation	Graphics were well-designed, clearly and legibly labeled, and informative to the investigation

	4 points	7 points	10 points
Results/Analysis (worth double)	Analysis showed many and/or severe deficiencies in clarity, specificity, design, or accuracy	Analysis fell somewhat short in clarity, specificity, design, or accuracy	Analysis of plots and statistics was thoroughly explained, appropriate to the investigation, and accurately interpreted
Code	Written code applies few or none of the coding and graphical design concepts from the course	Written code applies some of the coding and graphical design concepts from the course	Written code represents a clear and ambitious attempt by the student to apply both coding and graphical design concepts from the course
Notebook	Code was poorly commented or did not meet several of the requirements	Code comments could have been improved or one of the requirements for submission was not met	Code is appropriately commented, runs linearly, and was packaged and submitted appropriately
Slide Design	Slides were unreadable, extremely text heavy, or submitted in the wrong format	Slides were somewhat deficient in readability or were somewhat text heavy	Slides were well designed and submitted with an appropriately clear and readable format (including axis/plot labels!) with minimal text
Presentation Skills	The presentation as a whole had many and/or severe deficiencies in volume, appropriateness, pace, thoroughness, etc.	The presentation as a whole had some deficiencies in volume, appropriateness, pace, thoroughness, etc.	The presentation as a whole was practiced, thorough, and given at an appropriate pace and volume.
Presentation Content	Description of the context and results of the investigation was unclear or insufficient or was not accessible to peers	Description of the context and results of the investigation had some deficiencies in clarity or thoroughness or was not always accessible to peers	Student clearly and thoroughly explained the context and nature of their investigation and its results, including all necessary statistical, astronomical, and computational background necessary to interpret their results

Total: _____/100