

Homework-03

Ada Chibueze

2024-05-30

Problem 1: Multiple linear regression: model selection and construction (52 points)

Problem 1a: Make a table or list of all the models from class and the last one you constructed on your own. Write a caption for your table. (8 points)

Table 1. Linear Models. The table above shows the listed linear models and associated predictor values used for their construction. The column header “Model” assigns a specific name to each linear model, while the “Predictor Variable” header outlines each specific predictor variable used for model construction. The “+” symbols designates that the specific variable was used in its construction while the blank spaces signifies that specific variable was not used.

Problem 1b: Write a 5-6 sentence “statistical methods” section. (8 points)

To examine the influence of species (categorical), water treatment (categorical), and/or specific leaf area (SLA, numeric) on total biomass, I constructed 5 individual multiple linear models testing out various combinations of these predictor variables (refer to Problem 1a). To determine the model that best described the influence of the listed predictor variables on total biomass, I first evaluated each models individual AIC score along with its delta AIC score, to pick a model that had of balance of complexity and interpretation. Afterwards, I then checked to see which model adhered best to the assumption of a linear model (residuals are homoscedastic, residual are normally distributed, outliers affecting final model) to move forward with my final decision. To evaluate linear model assumptions visually I checked for constant variance of residuals by checking for a straight line and evenly distributed residual points on the Residuals versus Fitted and Scale-Location plots, made sure that the residuals followed along reference line shown on the QQ Residuals plot, and check to make sure no outlines fell outside of the Cook’s Line on Residual versus Leverage plot.

Problem 1c. Make a visualization of the model predictions with underlying data for your “best” model. (20 points)

Problem 2. Affective visualization (24 points)

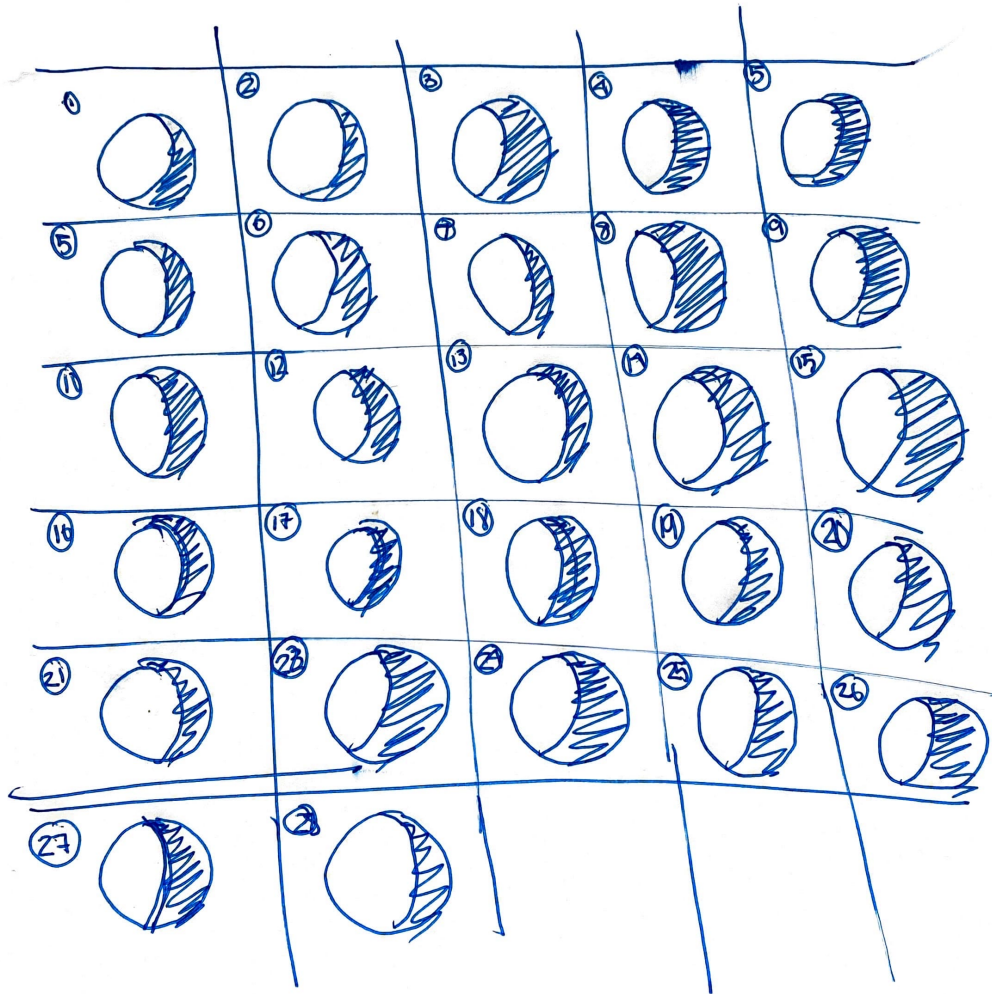
Problem 2a: Describe in words what an affective visualization could look like for your personal data (3-5 sentences). (2 points)

My personal data collection revolves around tracking sleep data, including the number of hours slept, time woken up, screen time before bed, etc. Affective visualization should thematically represent common symbols associated with sleep and nighttime to communicate my data collection to others. From this, I started to think about the moon and its various phases and how I could use them to represent some aspect of my data. Given this, each moon phase could represent an associated value of my sleep quality for the night.

Problem 2b: Create a sketch (on paper) of your idea. (2 points)

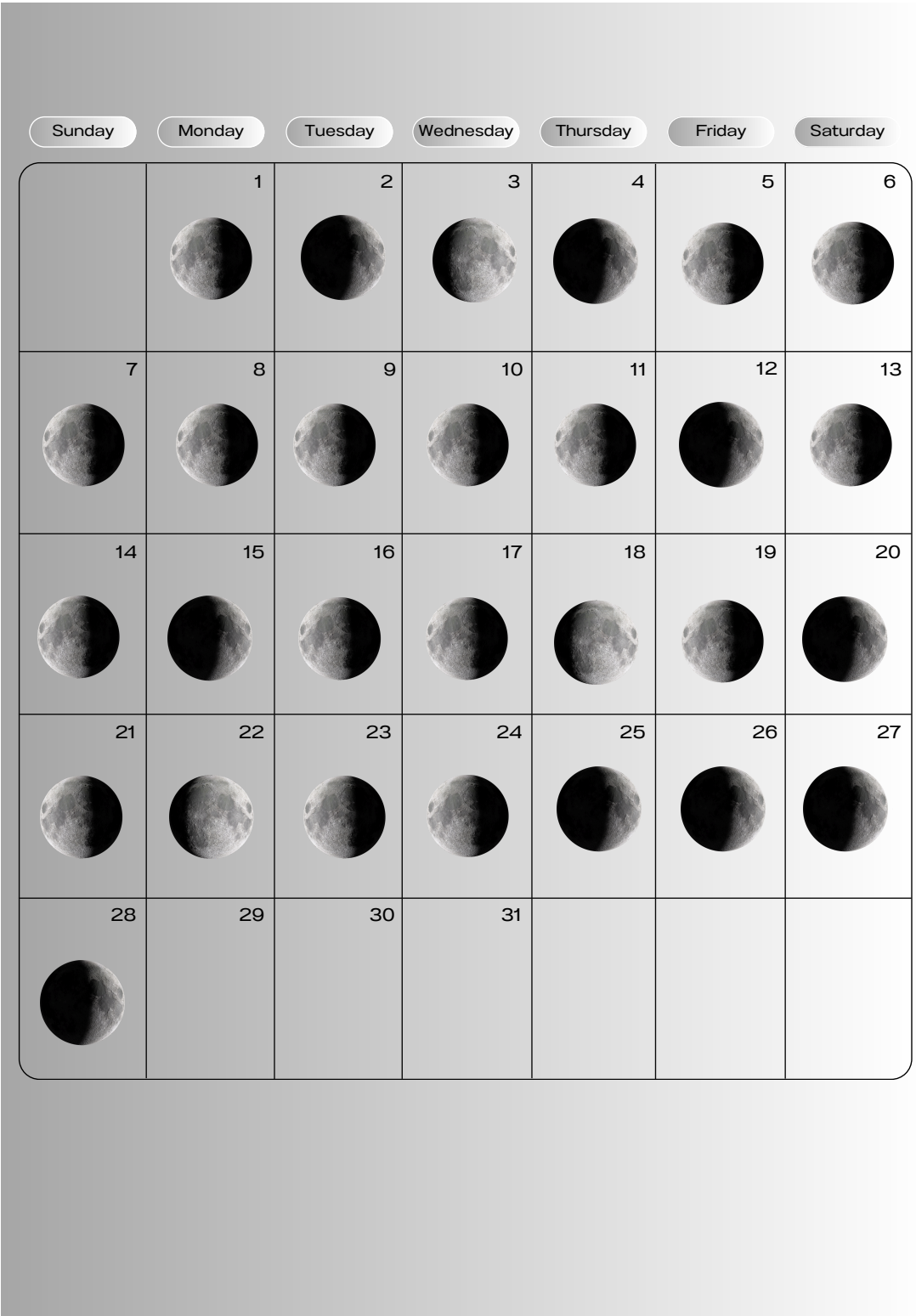
```
knitr::include_graphics("hw_figures/sleep_quality.pdf")
```

Sleep Quality Variation In May



Problem 2c: Make a draft of your visualization. (12 points)

```
knitr::include_graphics("hw_figures/sleep_visualization.pdf")
```



Problem 2d: Write an artist statement. (8 points)

In the piece I am showing how my sleep quality varies over the month of month of May. As the moon wanes it is indicative of a lower quality of sleeps, while as it waxes it represents my sleep quality increasing. I knew that I want to draw upon figures representative to sleep and night time hence the decision to use the moon, however for formatting I was inspired my some of the artistic pieces presented in Stefanie Posavec and Giorgia Lupi's Dear Data project. The form of my work is a visual display I made in Canvas. To make this visual I found different images of the moon online to represent different sleep levels, and then inserted them into the calendar I made to play on the idea of time and changes in sleep quality throughtout the month of May.

Problem 3. Statistical critique (36 points)

Problem 3a: Revisit and summarize (6 points)

**The authors used linear regression models to examine the relationship between housing vacancy and urban greening in Toledo, Ohio. They constructed separate linear regression models for the study dates of 1980, 2000, and 2004 to evaluate the relationship between mean NVDI vales and vacancy rates in Toledo Ohio. The authors also made a linear regression model to examine the influence of vacant housing units in Toledo on the amount of overgrown lawns/bumping (a sign of blight)

Problem 3b: Visual clarity (10 points)