1. What is design in software development?

Design in software development refers to the step in which you lay out the architectural decisions necessary to complete the process adequately. This step considers the requirements laid out for the project and then the team attempts to anticipate the challenges imposed by requirements. This anticipation allows the team to make design decisions early and modifications on the fly. The requirements will probably never stay the same throughout the project, but ensuring that the team starts out on the same page while designing is important.

1. List two principles of agile methods and describe them using specific examples that explain the principle’s meaning.

Two principles: Working Software over extensive documentation: meaning you don’t have to have extensive documentation in a time crunch, just enough to know whats going on and working software. It eliminates the need to have everything extensively planned out step by step and saves time, because you’re going to run into a problem which requires you to change your plan anyways. Responding to change over sticking with a plan: This principle builds off of the previous one that I mentioned, when it hits the fan in development you’ve got to be prepared to pivot. There is no point in crying over spilled milk when you realize that all of your planning was based upon assumptions that are no longer true. You adapt the plan and keep going.

1. What is “EDA” and why “EDA” is fundamentally a creative process?

EDA stands for Exploratory Data Analysis. The creativity of this process lies in discovering the trends of the data after beginning to analyze it. You enter the process with no specific aim to prove, you’re simply looking for trends and following your curiosity.

1. What is the difference between value and observation?

The difference between value and observation lies in the cases. Each response in a sample/dataset is an observation, whereas a value is more like a variable. It’s a property of the observation, a number or title attached to it.

1. In what ways can an outlier be helpful?

Outliers can illustrate the spread of a dataset well and display the fact that the data is not absolute, exceptions do occur. Outliers are also great at emphasizing the skew of a dataset.

1. Read about [filter](https://www.rdocumentation.org/packages/dplyr/versions/0.7.8/topics/filter) and provide examples of a multiple criteria and multiple arguments statements using your dataset. *Each student contributes with their example.*

filter(av\_tib, AveragePrice > 1, region == "Albany")

filter(av\_tib, region == "atlanta" | region == "Boise")

1. Create a histogram using your data, and explain your plot. Insert the image in the answer. *Each student contributes with their example.*

Chart

Description automatically generated This histogram depicts how many cases were taken from each year.

1. What is the difference between getwd() and setwd()?

Getwd() gets your working directory, where as setwd() changes your working directory to wherever you want it to be. Which allows you to access files with a relative path as opposed to an absolute one.

1. What does str() do?

According to [geeksforgeeks.org](https://www.geeksforgeeks.org/display-the-internal-structure-of-an-object-in-r-programming-str-function/) str() provides a compact summary of the data you give it.

1. Choose one of the following data visualizations projects and perform a what-why-how analysis (at least two charts) using the taxonomy presented by Munzner. *Each student contributes with their example.*

Top 100 golf courses in the US

What: Tables(items and attributes) ie. golf course rankings, names, location details, public/private. Geometry as well, to display locations on the map.

Why: Consume(discover, present, enjoy) & Query(compare, summarize). Targets: all data(trends), attributes(one). This vis takes the data and explores the rankings from two different golf magazines over time in order to display the top golf courses, while showing attributes such as public/private status. It also compares public vs private rankings over time. It seems that the private courses have faired better over time, especially with golf mag.

How: Encode(Arrange(use), Map(color)), Manipulate(select), Facet(superimpose), Reduce(embed). The data is display primarily on a map. The course’s status as to public vs private is communicated through color. You can select data from the map and discover its ranking and details (reduce(embed)). There is also a line chart in which they superimpose multiple ranking sources so as to compare their trends over time.