Part 1: Review Questions

1. The Cancer Genome Atlas (TCGA) is a project that looks at genomic data from different cancers. The project collects information on all of the genomic alterations through gene mutations, copy number variations, and changes to expression.

2. In terms of strengths, TCGA is a large dataset in being the largest dataset for cancer genomics. It is also open access meaning it can be accessed and analyzed in many different ways. There is a wide range of data including DNA sequencing, RNA sequencing, epigenetic data, patient demographics, survival data, and information on clinical history. In terms of weaknesses, some cancer types have more data than other cancer types. It is impossible to do an equal analysis across all cancer types. Additionally, not all patients have all of the clinical history recorded with many patients having the data recorded on older machinery that can be less accurate than the equipment used on patients in the later years.

3. git status, git add "file_name", git commit -m "Commit message", git push

4. if (!require (package)){install.packages("packages")} library(package)

5.
```
if (!require("BiocManager", quietly = TRUE))
    install.packages("BiocManager")
BiocManager::install(version = "3.16")
```

6. Boolean indexing allows you to filter out a dataset based on a certain set of conditions. This helps take only a specific section of the dataset that you would like to analyze. For example, patients who have used a specific type of chemotherapy or are above a certain age.

7.