Health Data Science: Homework #2
Karis Van Noord

1. What is the difference between a spreadsheet and a database?
    a. A spreadsheet is used to store and organize data in rows and columns. Within each cell a formula, function, or value can be inputted to represent the data. Spreadsheet software often includes tools used to sort, filter, and summarize data in a variety of ways. In class we experimented with some of these tools including Pivot Tables and Data Analysis. Some common platforms to utilize spreadsheets are Microsoft Excel, Apple Numbers, or Google Sheets. A database is an effective way to organize and store health data. Many databases make it easy to retrieve and organize data in a way that helps public health researchers, clinicians, and patient care. Databases normalize data, meaning they reduce anomalies and eliminate mistakes in data collection.
2. What are the advantages and disadvantages of using a spreadsheet for health data management?
    a. Although spreadsheets can be incredibly effective to store and organize data, they have unique file formats which are not always compatible with other software systems such as R or Python. There is a file form that is compatible with all softwares, called a CSV, however CSV's cannot contain cell formatting, formulas, charts, or other images that XLSX supports. In addition, the proper HIPAA regulations can be more difficult to implement using spreadsheets to store data rather than a secure database.
3. What are the advantages and the disadvantages of using a database for health data management?
    a. Databases can include many different kinds of information such as electronic health records, insurance claims, pharmaceutical research data, patient demographics and more. Databases also reduce costs and increase efficiency in patient care. Databases also normalize data, which is essentially cleaning up the collection of data and scanning for mistakes in the data input. Data normalization follows several "normal forms", which ensure that there is a unique identifier for each record in the table, each column contains a single value, and values stored in a column should be of the same domain. Although the use of databases is effective, it may require ongoing maintenance and updates. In addition, the systems may be prone to technical problems, or the systems may not be compatible with all devices and softwares. Some databases may even pose a threat to data security and privacy.
4. What are some examples of tasks that can be performed using a spreadsheet?

    a. Spreadsheets can be used to keep track of data and organize data. Additionally, spreadsheets can be used for financial analysis, accounting, time management, and task management.
5. What are some examples of tasks that can be performed using a database?
    a. Databases can keep track of many different kinds of information while also ensuring the correct individuals have access to the data. All the data that is collected and inputted into the database can be gathered and organized into one place to make data analysis easier and more efficient. Data is also proven to be more accurate in databases, because it is less likely to make mistakes when inputting data.
6. What are some examples of tasks that can be performed using both a spreadsheet and a database?
    a. While using a database and a spreadsheet one can perform extensive research and data analysis using the advantages of both softwares. Although both databases and spreadsheets have flaws, using the softwares together allows researchers to handle data and information of all kinds and sizes. Using both softwares, data integrity will be ensured and access to the data will be permitted to multiple individuals at once.
7. What are some examples of tasks that can be performed using neither a spreadsheet nor a database?
    a. Using neither a database nor a spreadsheet, researchers can still conduct descriptive, exploratory, predictive and quantitative/qualitative data analysis, but on a much smaller scale. Using the softwares is advantageous because it allows for the analysis of a much larger set of data.
8. What are some examples of tasks that can be performed using a spreadsheet but not a database?
    a. Using a spreadsheet, one can perform in-depth analysis, including calculations and specific formulas on small volumes of data. Spreadsheets are much more simple to use and have user-friendly features whereas databases require background knowledge about the design of databases and SQL.
9. What are some examples of tasks that can be performed using a database but not a spreadsheet?
    a. Using a database, one can handle much more complex structures of data, such as tables and relations. In addition, the integrity of data is ensured, while using a spreadsheet, there are not always measures to ensure data integrity.