Lauren Berryman

Prof. Minkoff

Data Diary

5 March 2022

**Final Project Data Diary**

**Graph 1**

**(corresponds with data sheet 2)**

1. Go to [CDC Wonder cancer database](https://wonder.cdc.gov/cancer.html) to find public information data on cancer mortality
2. Under “Cancer Mortality 1999-2018,” select “data request”
3. Under section 1…
   1. Turn setting on to group results by “year”
   2. Select “age adjusted rates”
4. Under section 5…
   1. Select “export results”
5. Click “send” at bottom of webpage
6. On Excel, click “file,” “open” and choose .txt file that was exported from CDC Wonder
7. When importing…
   1. Make sure “delimited” is selected and click “next”
   2. Make sure “tab” is selected and click “next”
   3. Make sure “general” is selected and click “finish”
8. Cleaning data…
   1. Copy and paste the data onto a new worksheet on the same file
   2. Delete “notes” column
   3. Change headings to “year,” “year\_code,” “deaths,” “population” and “age\_adjusted\_rate”
9. Select cleaned data with headings
10. Analyzing data…
    1. Select “data,” “summarize with pivot table,” “new worksheet” and “okay”
    2. Drag “year” to rows
    3. Drag “age\_adjusted\_rate” to values
       1. Click the “i” next to “age\_adjusted\_rate,” select “average” and click “okay”
11. To calculate age-adjusted rate percent change from 1999 to 2018, I used the values 200.7 (in 1999) and 149.2 (from 2018) and used the percent chance formula: =(new\_value-old\_value)/old\_value
    1. The number comes out to be -0.257, meaning there was a 25% decline in age-adjusted cancer mortality rate
12. Save file and change file format from .txt to .xlsx
13. I uploaded this cleaned data onto Datawrapper to depict this decline with a line graph
14. By completing these steps, I learned **how much the mortality rate decreased over 1999 to 2018**, the years CDC has data.

**Graph 2, 3 and 4**

**(corresponds with data sheet 1)**

1. Go to [CDC Wonder cancer database](https://wonder.cdc.gov/cancer.html) to find public information data on cancer mortality
2. Under “Cancer Mortality 1999-2018,” select “data request”
3. Under section 1…
   1. Turn setting on to group results by “cancer sites” and “race”
   2. Select “age adjusted rates”
4. Under section 3…
   1. Select “2018” as the year
5. Under section 4…
   1. Select “colon excluding rectum,” “female breast” and “prostate”
6. Under section 5…
   1. Select “export results”
7. Click “send” at bottom of webpage
8. On Excel, click “file,” “open” and choose .txt file that was exported from CDC Wonder
9. When importing…
   1. Make sure “delimited” is selected and click “next”
   2. Make sure “tab” is selected and click “next”
   3. Make sure “general” is selected and click “finish”
10. Cleaning data…
    1. Copy and paste the data onto a new worksheet on the same file
    2. Delete empty “notes” column
    3. Change headings to “cancer\_sites,” “cancer\_sites\_code,” “race,” “race\_code,” “deaths,” “population” and “age\_adjusted\_rate”
11. Select cleaned data with headings
12. Analyzing data…
    1. Select “data,” “summarize with pivot table,” “new worksheet” and “okay”
    2. Drag “cancer\_sites” to columns
    3. Drag “race” to rows
    4. Drag “age\_adjusted\_rate” to values
       1. Click the “i” next to “age\_adjusted\_rate,” select “average” and click “okay”
13. Save file and change file format from .txt to .xlsx
14. I uploaded this cleaned data onto Datawrapper to how age-adjusted mortality rate for colon, breast and prostate cancers varies by race with three separate bar graphs (by cancer site)
15. By completing these steps, I learned **how people of different races were impacted by colon cancer**, **how they were impacted by breast cancer** and **how they were impacted by prostate cancer in 2018**, the most recent year the CDC has data.

**Context in Story**

**(corresponds with data sheet 3)**

1. To confidently say that ***each*** year from 1999 to 2018, Black Americans had a higher age-adjusted mortality rate for colon, breast and prostate cancers, I analyzed data further.
2. Go to [CDC Wonder cancer database](https://wonder.cdc.gov/cancer.html) to find public information data on cancer mortality
3. Under “Cancer Mortality 1999-2018,” select “data request”
4. Under section 1…
   1. Turn setting on to group results by “cancer sites,” “race” and “year”
   2. Select “age adjusted rates”
5. Under section 4…
   1. Select “colon excluding rectum,” “female breast” and “prostate”
6. Under section 5…
   1. Select “export results”
7. Click “send” at bottom of webpage
8. On Excel, click “file,” “open” and choose .txt file that was exported from CDC Wonder, click “open”
9. When importing…
   1. Make sure “delimited” is selected and click “next”
   2. Make sure “tab” is selected and click “next”
   3. Make sure “general” is selected and click “finish”
10. Cleaning data…
    1. Copy and paste the data onto a new worksheet on the same file
    2. Delete empty “notes” column
    3. Change headings to “cancer\_sites,” “cancer\_sites\_code,” “race,” “race\_code,” “year,” “year\_code,” “deaths,” “population” and “age\_adjusted\_rate”
11. Select cleaned data with headings
12. Analyzing colon cancer data…
    1. Select all “data” on cleaned data sheet, “summarize with pivot table,” “new worksheet” and “okay”
    2. Drag “cancer\_sites” to filters
    3. Drag “year” and “race” to rows
    4. Drag “age\_adjusted\_rate” to values
       1. Click the “i” next to “age\_adjusted\_rate,” select “max” and click “okay”
    5. On the pivot table at B1 where it says “(All),” click the drop down button and filter only by “colon excluding rectum”
13. Analyzing breast cancer data…
    1. Select all “data” on cleaned data sheet, “summarize with pivot table,” “new worksheet” and “okay”
    2. Drag “cancer\_sites” to filters
    3. Drag “year” and “race” to rows
    4. Drag “age\_adjusted\_rate” to values
       1. Click the “i” next to “age\_adjusted\_rate,” select “max” and click “okay”
    5. On the pivot table at B1 where it says “(All),” click the drop down button and filter only by “female breast”
14. Analyzing prostate cancer data…
    1. Select all “data” on cleaned data sheet, “summarize with pivot table,” “new worksheet” and “okay”
    2. Drag “cancer\_sites” to filters
    3. Drag “year” and “race” to rows
    4. Drag “age\_adjusted\_rate” to values
       1. Click the “i” next to “age\_adjusted\_rate,” select “max” ” and click “okay”
    5. On the pivot table at B1 where it says “(All),” click the drop down button and filter only by “prostate”
15. Save file and change file format from .txt to .xlsx
16. By completing these steps, I learned **each year from 1999 to 2018,** **Black Americans had the highest mortality rates for colon cancer**, **breast cancer** and **prostate cancer compared to people of other races**.

**Outside Data Facts**

* **More than 1.7 million Americans were diagnosed with cancer in 2018**, according to the [Centers for Disease Control and Prevention](https://wonder.cdc.gov/cancer.html).
* The drug, also called Gleevec, improved outcomes for adults with chronic myelogenous leukemia, a blood cancer. **The drug more than quadrupled the survival rate from 22% to 90%**, according to the [American Cancer Society](https://www.cancer.org/cancer/chronic-myeloid-leukemia/detection-diagnosis-staging/survival-rates.html).
* From January to July 2020, **nearly 10 million Americans missed routine cancer screenings compared to data from 2019**, according to a [report](https://www.aacr.org/about-the-aacr/newsroom/news-releases/aacr-releases-report-outlining-impact-of-covid-19-pandemic-on-cancer-research-and-patient-care/) by the American Association for Cancer Research released last February.