ChanJacob_README

Meilin Chan and Natasha Jacob

4/6/2022

<Instructions: copy and paste this template into your project README file (found in the parent folder of the repository). Fill in relevant information as requested.>

<General notes: add as much information as is relevant for your repository. Some overarching guidelines are provided, but feel free to expand on these guidelines.> <More resources found here: https://www.dataone.org/all-best-practices> <Delete the text inside the brackets when formatting your file.>

Summary

The data within this repository includes a shapefile to conduct any spatial analysis, a raw data file which includes public health and population distribution data per NC county, and a data file which includes the spatial locations of hog farms within NC. The purpose of this repository is to examine if there are higher levels of health challenges and disparities within NC with high concentrations of hog farms compared to other NC counties, as well as observe any racial and economic inequities within communities where hog farms are mainly located. We will compare racial, health, etc. distributions within North Carolina compared to counties with the highest concentration of hog farms.

Investigators

Meilin Chan
meilin.chan@duke.edu

Natasha Jacob
natasha.jacob@duke.edu

Keywords

- hog farms
- environmental racism
- environmental health
- health disparities

Database Information

- County Health Data 2021.xlsx (Raw folder)
    - Source: [NCIOM Data] https://nciom.org/nc-health-data/map/?msclkid=9e415084b5bb11ec86c6f367b7f93e6c

    - accessed: April 6, 2022
- County Health Data 2021.csv (Process folder)

– Source: wrangled County Health Data 2021.xlsx to csv format
- cb_2019_us_county_20m.shp (Raw folder)
    – Source: Environmental_Data_Analytics_2022 Class GitHub Repository (Folder: Data/Spatial)

Folder structure, file formats, and naming conventions

- Folders *Data = stores all data used in the project

    – Raw = raw data, unwrangled
    – Processed = wrangled data
    – Spatial = spatial analysis data
    – README = folder containing the template and project-specific version of README which provides background and overview of the repository and purpose of the project.
    – Spatial Images = contains saved images of spatial analysis results

- Types of files *.csv = for wrangling and creation of dataframes for ggplot visualizations and other analyses* .shp = for spatial analyses

- Naming of files

    – Raw data file names are kept as the original file names when pulled from the source data bases.
    – Processed data file names include a short description of what the files contain

Metadata

Raw Folder

We wrangled the original County Health Data 2021.xlsx into County Health Data 2021.csv file for easier access with manipulating the dataset.The County Health Data 2021.csv file has 62 rows and 109 columns. The class of all the columns is a factor. The dataframe contains descriptions for each row (Bold.Indicator.Name.Text) in the Indicator.Descriptor.Text and Source.Label columns.

```
##                               X
##                               :16
##  Demographics                 :10
##  Access to Care               : 7
##  Maternal & Child Health      : 6
##  Tobacco & Substance Use      : 5
##  Healthy Eating, Active Living: 4
##  (Other)                      :14
##                               Bold.Indicator.Name.Text
##                                      : 8
##  Racial Disparity in Infant Mortality Rate: 2
##  Racial Disparity in Life Expectancy      : 2
##  Adult Obesity                            : 1
##  Adult Smoking                            : 1
##  African American                         : 1
##  (Other)                                  :47
##                                                              Indicator.Descriptor.Text
##                                                                     : 9
##  % of 10-17 years olds who are overweight or obese                  : 1
##  % of adults age 18 to 64 years without health insurance            : 1
##  % of adults consuming fruits, vegetables, or beans five or more times per day    : 1
##  % of adults who are current smokers                                : 1
##  % of adults who have had permanent teeth removed due to tooth decay or gum disease: 1
##  (Other)                                                            :48
##                                                                                    Source
```

```
##  NC State Center for Health Statistics
##
##  US Census - Population Estimates Program
##  CDC
##  CDC - Behavioral Risk Factors Survey
##  North Carolina Health Preofessions Data System - Cecil G. Sheps Center for Health Services Research
##  (Other)
##                                                  Source.Link.s.
##  https://www.census.gov/quickfacts/fact/table/NC/PST045219:11
##                                                           : 8
##  http://www.ncpublicschools.org/accountability/reporting/ : 2
##  http://www.schs.state.nc.us/data/lifexpectancy/          : 2
##  https://nchealthworkforce.sirs.unc.edu/                  : 2
##  https://nchealthworkforce.unc.edu/supply/                : 2
##  (Other)                                                  :35
##    X.1                  NC              Counties      X.2          X.3
##  Mode:logical        :10                   : 7           : 7          : 7
##  NA's:62       0.1%      : 1   22.9%       : 2   N/A    : 2   -      : 3
##                1.6%      : 1   N/A         : 2   -      : 1   N/A    : 3
##                10,488,084: 1   *Central: 25.2%: 1   0      : 1   1.7    : 2
##                10.1%     : 1   0.0%        : 1   0.10%  : 1   17.0%  : 2
##                11.8      : 1   0.1%        : 1   0.5%   : 1   **     : 1
##                (Other)   :47   (Other)     :48   (Other):49   (Other):44
##      X.4          X.5          X.6          X.7          X.8          X.9
##         : 7          : 7          : 7          : 7          : 7          : 7
##  -      : 2   N/A    : 3   -      : 3   3.8    : 2   N/A    : 2   -      : 2
##  1.2    : 2   -      : 2   0.6%   : 2   8.0%   : 2   -      : 1   N/A    : 2
##  16.6%  : 2   **     : 1   N/A    : 2   N/A    : 2   **     : 1   0      : 1
##  N/A    : 2   0.10%  : 1   **     : 1   -      : 1   0      : 1   0.3%   : 1
##  **     : 1   0.4%   : 1   0.1%   : 1   0.1%   : 1   0.0    : 1   1      : 1
##  (Other):46   (Other):47   (Other):46   (Other):47   (Other):49   (Other):48
##      X.10         X.11         X.12         X.13         X.14         X.15
##         : 7          : 7          : 7          : 7          : 7          : 7
##  -      : 2   5.2%   : 2   18.4%  : 2   11.1%  : 2   2      : 2   -      : 3
##  10.2%  : 2   N/A    : 2   N/A    : 2   13.4%  : 2   4.6%   : 2   0      : 2
##  N/A    : 2   0.2%   : 1   -      : 1   N/A    : 2   N/A    : 2   0.9    : 2
##  0.1%   : 1   0.5%   : 1   0.7%   : 1   0.1%   : 1   -      : 1   3.0%   : 2
##  0.7%   : 1   1.4%   : 1   0.9%   : 1   0.7%   : 1   0.1    : 1   N/A    : 2
##  (Other):47   (Other):48   (Other):48   (Other):47   (Other):47   (Other):44
##      X.16                  X.17         X.18         X.19         X.20
##         : 7                  : 7          : 7          : 7          : 7
##  5.6    : 2   -             : 3   21.5%  : 2   3      : 2   -      : 2
##  N/A    : 2   N/A           : 2   N/A    : 2   8.2    : 2   1.6%   : 2
##  0.2%   : 1   **            : 1   0.1%   : 1   8.7%   : 2   2      : 2
##  0.6%   : 1   *Central: 25.2%: 1   0.6%   : 1   N/A    : 2   2.4    : 2
##  1.3%   : 1   0             : 1   1.5    : 1   -      : 1   76     : 2
##  (Other):48   (Other)       :47   (Other):48   (Other):46   (Other):45
##      X.21         X.22         X.23         X.24         X.25
##         : 7          : 7          : 7          : 7          : 7
##  -      : 2   -      : 4   19.0%  : 2   5.6%   : 2   N/A    : 2
##  N/A    : 2   20.4%  : 2   N/A    : 2   N/A    : 2   0.2%   : 1
##  **     : 1   3.4    : 2   -      : 1   -      : 1   0.7%   : 1
##  0.2%   : 1   76.5   : 2   0.4    : 1   0.1%   : 1   102,139: 1
##  0.5%   : 1   N/A    : 2   0.4%   : 1   0.4    : 1   11.7%  : 1
```

```
##  (Other):48    (Other):43    (Other):48    (Other):48    (Other):49
##               X.26          X.27          X.28          X.29          X.30
##                  : 7           : 7           : 7           : 7           : 7
## 18.0%        : 2   -       : 2   -       : 3  N/A    : 2  -       : 4
## N/A          : 2  8.8%    : 2  N/A    : 2  0.1%    : 1  10.9%  : 2
## *Central: 25.2%: 1  N/A     : 2  0.1%    : 1  0.7     : 1  21.7%  : 2
## 0.4%         : 1  0       : 1  0.6%    : 1  0.8%    : 1  3.5     : 2
## 1            : 1  0.1%    : 1  0.9%    : 1  1.6%    : 1  N/A     : 2
## (Other)      :48  (Other):47  (Other):47  (Other):49  (Other):43
##     X.31              X.32          X.33          X.34
##         : 7              : 7           : 7           : 7
## N/A    : 2  N/A            : 2  11.6%  : 2  N/A     : 2
## -      : 1  *Central: 25.2%: 1  21.0%  : 2  0.1%    : 1
## 0.4%   : 1  0.1%           : 1  N/A    : 2  0.9%    : 1
## 0.5    : 1  0.9%           : 1  -      : 1  10.1%   : 1
## 0.7%   : 1  10.6           : 1  0.1%   : 1  10.5%   : 1
## (Other):49  (Other)        :49  (Other):47  (Other):49
##               X.35          X.36          X.37          X.38          X.39
##                  : 7           : 7           : 7           : 7           : 7
## 2            : 2  2       : 2  -       : 3  -       : 3  15.6    : 2
## N/A          : 2  26.0%   : 2  0       : 3  0.5%    : 2  2.4     : 2
## *Central: 25.2%: 1  N/A     : 2  N/A    : 3  7.9%    : 2  6.5     : 2
## 0.1          : 1  0.1%    : 1  12.7%   : 2  N/A     : 2  N/A     : 2
## 0.1%         : 1  0.6%    : 1  2.4%    : 2  **      : 1  0.1%    : 1
## (Other)      :48  (Other):47  (Other):42  (Other):45  (Other):46
##     X.40              X.41                  X.42                  X.43
##         : 7              : 7               : 7                   : 7
## N/A    : 3  16.0%          : 2  -            : 2  7.3              : 2
## -      : 2  N/A            : 2  N/A          : 2  N/A              : 2
## 1      : 2  *Central: 25.2%: 1  *Central: 25.2%: 1  -              : 1
## **     : 1  0.1%           : 1  0.1%          : 1  *Central: 25.2%: 1
## 0.3%   : 1  0.8%           : 1  0.8           : 1  0.2%           : 1
## (Other):46  (Other)        :48  (Other)       :48  (Other)        :48
##     X.44          X.45          X.46                  X.47          X.48
##         : 7           : 7           : 7               : 7           : 7
## -      : 2  -       : 2  2.5     : 2  -            : 2  -       : 5
## 1.4%   : 2  N/A     : 2  N/A    : 2  N/A            : 2  0       : 2
## 7.1    : 2  0.2%    : 1  **      : 1  *Central: 25.2%: 1  17.0%  : 2
## 77.1   : 2  0.7%    : 1  0.1%    : 1  0.4            : 1  N/A     : 2
## N/A    : 2  1.3%    : 1  0.8%    : 1  0.4%           : 1  **      : 1
## (Other):45  (Other):48  (Other):48  (Other)        :48  (Other):43
##     X.49          X.50          X.51          X.52                  X.53
##         : 7           : 7           : 7           : 7               : 7
## N/A    : 2  -       : 4  0.9%    : 2  -       : 4  14.2%          : 2
## 0.1%   : 1  18.0%   : 2  3       : 2  1       : 2  5.4            : 2
## 0.5%   : 1  N/A     : 2  N/A    : 2  75.9    : 2  N/A            : 2
## 1.3    : 1  1.1%    : 1  0.1%    : 1  N/A     : 2  *Central: 25.2%: 1
## 10.2%  : 1  12.0    : 1  0.5     : 1  **      : 1  0.2%           : 1
## (Other):49  (Other):45  (Other):47  (Other):44  (Other)        :47
##     X.54          X.55          X.56          X.57          X.58          X.59
##         : 7           : 7           : 7           : 7           : 7           : 7
## -      : 2  3       : 2  -       : 2  -       : 4  -       : 2  7.1     : 2
## 12.8%  : 2  N/A     : 2  77.9    : 2  0.5%    : 2  N/A     : 2  76.9    : 2
## N/A    : 2  -       : 1  N/A    : 2  14.7%   : 2  **      : 1  N/A     : 2
```

```
##  0.2%    : 1    0.1    : 1   >95.0% : 1    77.2   : 2    0.2%    : 1    -       : 1
##  0.6%    : 1    0.1%   : 1    0.1%   : 1    N/A    : 2    0.4     : 1    0.1%    : 1
##  (Other):47    (Other):48    (Other):47    (Other):43    (Other):48    (Other):47
##         X.60           X.61                  X.62                  X.63
##          : 7            : 7                   : 7                   : 7
##  29.0%    : 2    -       : 3    -              : 2    N/A            : 2
##  N/A      : 2    N/A     : 3    N/A            : 2    -              : 1
##  0.1%     : 1    8.5     : 2    **             : 1    *Central: 25.2%: 1
##  0.8%     : 1    0.2%    : 1    *Central: 25.2%: 1    0.2%           : 1
##  1,110,356: 1    0.7     : 1    0              : 1    1.0%           : 1
##  (Other)  :48    (Other):45    (Other)        :48    (Other)        :49
##               X.64           X.65           X.66           X.67
##                : 7            : 7            : 7            : 7
##  1.0%          : 2    5.5    : 2    -      : 4    N/A     : 2
##  75.2          : 2    N/A    : 2    1      : 2    0.3%    : 1
##  N/A           : 2    0.1%   : 1    N/A    : 2    0.9%    : 1
##  -             : 1    0.6%   : 1    **     : 1    10.3    : 1
##  *Central: 25.2%: 1    1.6%   : 1    0      : 1    10.8%   : 1
##  (Other)       :47    (Other):48    (Other):45    (Other):49
##               X.68           X.69           X.70           X.71           X.72
##                : 7            : 7            : 7            : 7            : 7
##  N/A           : 2    -       : 2    -       : 2    N/A     : 2    -       : 4
##  -             : 1    15.1%   : 2    N/A     : 2    -       : 1    0.5%    : 2
##  *Central: 25.2%: 1    76.5    : 2    0.1%    : 1    0.1%    : 1    N/A     : 2
##  0.1%          : 1    N/A     : 2    0.6%    : 1    0.7%    : 1    **      : 1
##  0.6%          : 1    >95.0% : 1    1       : 1    0.9%    : 1    0.7     : 1
##  (Other)       :49    (Other):46    (Other):48    (Other):49    (Other):45
##               X.73           X.74           X.75                  X.76
##                : 7            : 7            : 7                   : 7
##  -             : 2    11.1    : 2    -      : 3    -              : 2
##  10.1          : 2    16.4    : 2    N/A    : 2    N/A            : 2
##  2             : 2    N/A     : 2    0.4%   : 1    *Central: 25.2%: 1
##  N/A           : 2    0.1%    : 1    0.6%   : 1    0.1%           : 1
##  *Central: 25.2%: 1    0.5%    : 1    0.8%   : 1    0.5            : 1
##  (Other)       :46    (Other):47    (Other):47    (Other)        :48
##               X.77           X.78                  X.79           X.80
##                : 7            : 7                   : 7            : 7
##  N/A           : 2    24.7%   : 2    20.9%          : 2    N/A     : 2
##  -             : 1    N/A     : 2    N/A            : 2    0.1%    : 1
##  *Central: 25.2%: 1    0.2     : 1    *Central: 25.2%: 1    0.6%    : 1
##  0             : 1    0.2%    : 1    0.1%           : 1    1.2%    : 1
##  0.2%          : 1    0.7%    : 1    0.6%           : 1    1.8%    : 1
##  (Other)       :49    (Other):48    (Other)        :48    (Other):49
##       X.81           X.82                  X.83           X.84           X.85
##          : 7            : 7                   : 7            : 7            : 7
##  -       : 2    N/A     : 2    10.2           : 2    10.5    : 2    -       : 3
##  N/A     : 2    -       : 1    N/A            : 2    4.1     : 2    13.0%   : 2
##  0.1%    : 1    0.2     : 1    -              : 1    75.5    : 2    4.1%    : 2
##  0.4%    : 1    0.3%    : 1    *Central: 25.2%: 1    N/A     : 2    76.9    : 2
##  0.6%    : 1    0.6%    : 1    0.10%          : 1    0.3%    : 1    N/A     : 2
##  (Other):48    (Other):49    (Other)        :48    (Other):46    (Other):44
##       X.86           X.87           X.88           X.89           X.90
##          : 7            : 7            : 7            : 7            : 7
##  -       : 2    -       : 4    -       : 2    -       : 4    N/A     : 2
```

```
##   N/A    : 2   22.0% : 2   N/A    : 2   0      : 4   0.1%   : 1
##   0.1%   : 1   N/A   : 2   0.1%   : 1   21.0%  : 2   0.5    : 1
##   0.4    : 1   **    : 1   0.3    : 1   N/A    : 2   0.6%   : 1
##   0.7%   : 1   0.7%  : 1   0.5%   : 1   **     : 1   10.9   : 1
##   (Other):48   (Other):45  (Other):48  (Other):42  (Other):49
##                 X.91                X.92                X.93           X.94
##                     : 7                 : 7                 : 7             : 7
##   3.5             : 2   10.4%           : 2   -               : 3   -       : 4
##   N/A             : 2   12.0%           : 2   0.5             : 2   N/A     : 2
##   *Central: 25.2%: 1   3               : 2   N/A             : 2   **      : 1
##   0.1%            : 1   N/A             : 2   **              : 1   0       : 1
##   0.2             : 1   *Central: 25.2%: 1   *Central: 25.2%: 1   0.4%    : 1
##   (Other)         :48   (Other)         :46   (Other)         :46   (Other):46
##       X.95         X.96         X.97                X.98           X.99
##           : 7          : 7          : 7                 : 7             : 7
##   -       : 3   N/A   : 2   N/A    : 2   N/A             : 2   -       : 2
##   81.3    : 2   0.2%  : 1   -      : 1   *Central: 25.2%: 1   N/A     : 2
##   N/A     : 2   0.9%  : 1   0.1%   : 1   0.1%            : 1   0.1%    : 1
##   0.4%    : 1   1     : 1   0.4    : 1   0.6%            : 1   0.5%    : 1
##   1.2%    : 1   1.3%  : 1   0.4%   : 1   1               : 1   0.7%    : 1
##   (Other):46   (Other):49   (Other):49   (Other)         :49   (Other):48
##       X.100         X.101
##           : 7   2019     :19
##   -       : 3   2017      : 8
##   1.1%    : 3              : 7
##   0.3%    : 2   2018      : 7
##   78.8    : 2   2014-2018: 4
##   N/A     : 2   2013-2017: 3
##   (Other):43   (Other)  :14
##
##
## Data are reported for NC Association of Local Health Director Regions
## Life expectancy is the average number of additional years that someone at a given age would be expe
## County-level data not available
## Rates based on fewer than 10 cases are suppressed
## *Technical Note: Rates based on small numbers (fewer than 10 deaths) are unstable & should be interp
## (Other)

## [1]  62 109

## [1] "data.frame"
```

Processed

The projectNCcounties_HealthData.csv file is our wrangled dataframe in which we conducted analysis on. We wrangled all the columns with filtered into our processed dataset into the numerical class.

```
##           X          Population       Child_Population_percentage
##   Beaufort:1   Min.   :    17557   Min.   :12.70
##   Bladen  :1   1st Qu.:    37573   1st Qu.:17.48
##   Duplin  :1   Median :    57459   Median :20.40
##   Durham  :1   Mean   :  1004715   Mean   :19.70
##   Greene  :1   3rd Qu.:    67375   3rd Qu.:22.18
##   Martin  :1   Max.   :10488084   Max.   :23.80
##   (Other) :6
```

```
## Elderly_Population_percentage White_Population_percentage
## Min.   :12.00                  Min.   : 34.70
## 1st Qu.:17.45                  1st Qu.: 58.83
## Median :20.55                  Median : 83.35
## Mean   :19.68                  Mean   :113.12
## 3rd Qu.:22.38                  3rd Qu.: 89.22
## Max.   :23.60                  Max.   :557.80
##
## Hispanic_Latinx_percentage African_American_percentage
## Min.   : 2.400             Min.   : 1.50
## 1st Qu.: 4.600             1st Qu.: 4.70
## Median : 6.850             Median :15.45
## Mean   : 7.617             Mean   :18.57
## 3rd Qu.: 8.825             3rd Qu.:27.10
## Max.   :23.000             Max.   :61.10
##
## American_Indian_percentage Reading_Proficiency_percentage
## Min.   :0.400              Min.   :24.60
## 1st Qu.:0.475              1st Qu.:49.58
## Median :0.700              Median :56.95
## Mean   :0.975              Mean   :54.39
## 3rd Qu.:1.025              3rd Qu.:62.05
## Max.   :3.300              Max.   :74.30
##
## Uninsured_Adults_percentage Medicaid_CHIP_enrolles_percentage
## Min.   :11.80               Min.   :10.50
## 1st Qu.:14.93               1st Qu.:21.35
## Median :15.50               Median :25.90
## Mean   :16.82               Mean   :25.83
## 3rd Qu.:17.52               3rd Qu.:29.95
## Max.   :25.70               Max.   :42.80
##
## Primary_Care_Physicians Low_Birthweight_percentage Infant_Mortality_Rate
## Min.   :2.200           Min.   : 7.900             Min.   : 4.000
## 1st Qu.:3.750           1st Qu.: 8.600             1st Qu.: 5.675
## Median :4.950           Median : 9.000             Median : 6.750
## Mean   :5.050           Mean   : 9.683             Mean   : 7.508
## 3rd Qu.:5.925           3rd Qu.:10.100             3rd Qu.: 9.200
## Max.   :8.500           Max.   :13.300             Max.   :12.600
##
## Cancer_Incidence Heart_Disease   Poverty_percentage Air_Pollution
## Min.   :407.9    Min.   :118.1   Min.   : 8.00      Min.   : 8.300
## 1st Qu.:433.8    1st Qu.:150.7   1st Qu.:14.35      1st Qu.: 9.025
## Median :454.4    Median :163.6   Median :15.80      Median : 9.850
## Mean   :453.6    Mean   :164.0   Mean   :16.74      Mean   : 9.683
## 3rd Qu.:464.8    3rd Qu.:171.2   3rd Qu.:19.23      3rd Qu.:10.250
## Max.   :529.6    Max.   :241.4   Max.   :25.80      Max.   :11.000
##
## [1] 12 18

## [1] "data.frame"
```

Column Name	Description - all per county
Population	Total population
Child_Population_percentage	% (per county)
Elderly_Population_percentage	% (per county)
White_Population_percentage	% (per county)
Hispanic_Latinx_percentage	% (per county)
African_American_percentage	% (per county)
American_Indian_percentage	% (per county)
Reading_Proficiency_percentage	% literacy rate per county
Uninsured_Adults_percentage	% adults w/out health insurance
Medicaid_CHIP_enrolles_percentage	% individ. w/ Medicaid or CHIP
Primary_Care_Physicians	Physicians per 10,000 pop.
Low_Birthweight_percentage	% of low birth weight occurrences
Infant_Mortality_Rate	% rate of infant mortality
Cancer_Incidence	Incidence rates per 100,000 pop.
Heart_Disease	Death rates per 100,000 pop.
Poverty_percentage	% of pop. living under poverty line
Air_Pollution	avg. daily density of PM (ug/m3)

Spatial

For the spatial analysis component of our project, we used the Cartographic Boundary shapefile (cb_2018_us_county_20m.shp) to map North Carolina and its counties. The cartographic boundary files are simplified representations of selected geographic areas from the Census Bureau's MAF/TIGER geographic database. These boundary files are specifically designed for small scale thematic mapping. We used 2018 data year shapefile with a resolution level of 20m. The shapefile contains information regarding geographic features of all US states. For our project, we were only interested in the counties located in North Carolina, hence, we filtered the shapefile to provide geographic information of the counties located in North Carolina. The attributes present in the county shapefile are the geographic coordinates, area of land and water, state and county FIPS codes, the American National Standards Institute (ANSI) codes, GEOIDs, names of the counties and legal/statistical area description codes.

```
library(sf)
```

```
## Linking to GEOS 3.9.1, GDAL 3.4.0, PROJ 8.1.1; sf_use_s2() is TRUE
```

```
NorthCarolina_sf <- st_read('../ChanJacob_ENV872_EDA_FinalProject/Data/Spatial/cb_2018_us_county_20m.shp
  filter(STATEFP == 37)
```

```
## Reading layer `cb_2018_us_county_20m' from data source
##   `/Users/natashajacob/Desktop/EDA872/Environmental_Data_Analytics_2022/EDA_Project/ChanJacob_ENV872_
##   using driver `ESRI Shapefile'
## Simple feature collection with 3220 features and 9 fields
## Geometry type: MULTIPOLYGON
## Dimension:     XY
## Bounding box:  xmin: -179.1743 ymin: 17.91377 xmax: 179.7739 ymax: 71.35256
## Geodetic CRS:  NAD83
```

```
colnames(NorthCarolina_sf)
```

```
##  [1] "STATEFP"  "COUNTYFP" "COUNTYNS" "AFFGEOID" "GEOID"    "NAME"
##  [7] "LSAD"     "ALAND"    "AWATER"   "geometry"
```

```
class(NorthCarolina_sf)
```

```
## [1] "sf"          "data.frame"
```

```
summary(NorthCarolina_sf)
```

```
##    STATEFP             COUNTYFP            COUNTYNS            AFFGEOID
## Length:100          Length:100          Length:100          Length:100
## Class :character    Class :character    Class :character    Class :character
## Mode  :character    Mode  :character    Mode  :character    Mode  :character
##
##
##
##    GEOID               NAME                LSAD                ALAND
## Length:100          Length:100          Length:100          Min.   :4.472e+08
## Class :character    Class :character    Class :character    1st Qu.:8.936e+08
## Mode  :character    Mode  :character    Mode  :character    Median :1.192e+09
##                                                             Mean   :1.259e+09
##                                                             3rd Qu.:1.501e+09
##                                                             Max.   :2.453e+09
##    AWATER                      geometry
## Min.   :4.453e+05    MULTIPOLYGON :100
## 1st Qu.:7.219e+06    epsg:4269    :  0
## Median :1.595e+07    +proj=long...:  0
## Mean   :1.347e+08
## 3rd Qu.:3.829e+07
## Max.   :3.001e+09
```

Column Name	Description
STATEFP	Federal Information Processing Standards (FIPS) code for North Carolina
COUNTYFP	FIPS code for each county in North Carolina
COUNTYNS	American National Standards Institute (ANSI) code for the county
AFFGEOID	American FactFinder geographic identifier
GEOID	A concatenation of current state FIPS code and county FIPS code
NAME	Names of counties in North Carolina
LSAD	Legal and statistical area description codes
ALAND	Current land area (in square meters)
AWATER	Current water area (in square meters)
geometry	Geographic coordinates of counties in North Carolina

The Hog_Farm_Locations_NC.csv file contains locations of Hog Farms present in North Carolina. The coordinates of Hog Farm Locations were obtained from Google Maps. Phantom Buster - a web application used for extracting data from websites was used to extract coordinates of Hog Farms from Google Maps. The coordinates were converted to a spatial features dataset inorder to map them with the counties dataset.

```
HogFarm_Locations <- read.csv("../ChanJacob_ENV872_EDA_FinalProject/Data/Spatial/Hog_Farm_Locations_NC1

#Selecting the required columns
HogFarm_Locations <- HogFarm_Locations %>%
  select(title, latitude, longitude) %>%
   na.omit(HogFarm_Locations)

#Converting the dataframe to a sf (spatial features) dataframe
HogFarm_Locations_sf <- HogFarm_Locations %>%
  st_as_sf(coords = c('longitude', 'latitude'),
```

```
        crs = 4269)
```

```
colnames(HogFarm_Locations_sf)
```

```
## [1] "title"    "geometry"
```

```
class(HogFarm_Locations_sf)
```

```
## [1] "sf"        "data.frame"
```

```
summary(HogFarm_Locations_sf)
```

```
##     title                geometry
##  Length:186      POINT       :186
##  Class :character  epsg:4269   :  0
##  Mode  :character  +proj=long...:  0
```

Column Name	Description
title	Names of Hog Farms
geometry	Geographic coordinates of the Hog Farms

<For each data file in the repository, describe the data contained in each column. Include the column name, a description of the information, the class of data, and any units associated with the data. Create a list or table for each data file.>

Scripts and code

No scripts or code are contained in our repository.

Quality assurance/quality control

We sourced our data from reputible sources such as Google Maps, the Environmental_Data_Analytics_2022 GitHub, and the North Carolina Institute of Medicine. It should be noted that some existing hog farms are not within the Google Maps database and as such were not included within the spatial analysis. To ensure high quality of consistancy in our data, we kept organized unwrangled data in a Raw folder, dataframes we had wrangled and saved in a Processed folder, and any data used with spatial analysis into a Spatial folder. We renamed our data files using names that would provide users a general idea of what the data contained. We never wrangled or edited the Raw data files, and agreed to only save wrangled data that would be used widely in our analyses. All our data and project files were saved and backed up to our GitHub repository which we pulled from at the beginning of each session, and pushed to at the end of each session.