Geospatial for Everyone

Nathaniel MacNell, Ian Buller

Enhancing your "Non-Spatial" Research with Geospatial Data

A free workshop for the 2025 NIH Research Festival presented by Ian Buller and Nat MacNell from DLH.

Contact us with any questions: - nathaniel.macnell@dlhcorp.com - ian.buller@dlhcorp.com

If you're reading a PDF copy, the interactive version can be found here: https://github.com/nathanielmacnell/nihworkshop/ - You'll need a free Google account to use the interactive version.

START HERE!

1. **Click the play button** below to run the setup (it may show up as [ ] until you mouse over it).
2. **Click the Table of contents** on the left to see sections.
3. **Keep this browser window open** (changing tabs is okay).

- If you're accessing this workshop in the future and you don't get a **check** after pressing the play button, try the **Installing from CRAN** section under **Further Reading** near the end of this document.

```
### <- Click the play button here to set up your notebook

# download the libraries and unpack
download.file('https://dlhcorp-my.sharepoint.com/:u:/p/nathaniel_macnell/EXnWcXbUQbNNhtKA-
system('tar -xzvf rlib.tar.gz')

# add library location to path
.libPaths(c("library", .libPaths()))

# load libraries
```

```r
library(sf)
library(ggplot2)
library(dplyr)
library(tigris)
```

```
Linking to GEOS 3.12.1, GDAL 3.8.4, PROJ 9.3.1; sf_use_s2() is TRUE


Attaching package: 'dplyr'


The following objects are masked from 'package:stats':

    filter, lag


The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union


To enable caching of data, set `options(tigris_use_cache = TRUE)`
in your R script or .Rprofile.
```

Background

Why use geospatial data?

We can increase the value of data we already have, by linking it to public-domain information using geographic identifiers. For example, a clinical research study about an off-label use of a drug could add:

- **Effect-measure modifiers** - are there contexts in which the drug is more effective?
- **Covariates** - how does the study population compare to the target population?
- **Confounders** - are there factors that could bias effect estimates?
- **Exclusions** - should some observations be removed from consideration?

We can also look at additional uses of geostatistical methods: - **Clustering** - are nearby observations similar, and does it matter? - **Aggregation** - can we understand at a problem at different levels? - **Imputation** - can we fill in missing values using spatial context?

Tools of the trade

There are a few languages with geospatial packages that are particularly helpful: - **R** - ***sf***, ***terra***, and ***ggplot2*** packages - **Python** - ***geopandas***, ***rasterio***, and ***matplotlib*** libraries

You can also use desktop GUI programs like ESRI's ArcGIS or the free equivalent QGIS, but these are much more helpful if you are creating or editing geospatial data, rather than analyzing it.

You'll come across many kinds of geospatial data formats, here are some of the most common:

Data type	Comments
.csv	The most basic type, literally just comma-separated values with latitude and longitude coordinates.
shapefile	The legacy geospatial data type, originally designed by ESRI in an era where computers worked very differently than they do today. You need multiple files with the same name that store different parts of the data.
.gdb	Geodatabase, ESRI's approach to bundling shapefiles.
.kml	A more modern file format from Google. It's XML-based, so you can open it up in a text editor and it's widely compatible.
GeoJSON	Another modern format, similar to kml but based on JSON instead of XML.
.GeoTIFF	An plain image with extra geospatial data embedded in it. These can often be inspected in a standard image viewer.

Today we're starting with R and .csv because they are the easiest to use and give us the best view of what's going on "behind the curtain".

Basics

Colab

This document is a Jupyter notebook hosted on Colab, a free service for running test code on the cloud (thanks, Google!). - The menu on the left has a **Table of Contents** for navigation. - You can also see the (temporary) **Files** in your virtual workspace.

Jupyter

Jupyter Notebooks help you run code on a laptop, the cloud, or the NIH Biowulf HPC. Here's the basics: * **Text cells**: Double-click to edit. * Text cells use markdown for formatting. * **Code cells**: Click the **Play** button to run them. * **Move cells**: Single click and use the pop-up menu .

You can collapse a section of the notebook by clicking the arrow next to the header.

R

- R uses a sequence of expressions with no line-end character
- Lines starting with # are comments and are not executed.
- The core concept is the function, which takes arguments (seperated by commas) and returns one result.

```
# example: add some numbers
sum(1,2,3,4,5)
```

15

- You can also use mathematical operators as expected

```
1+2+3+4+5
```

15

- To store results, use an equation or assignment arrow. Otherwise, the results will be printed out immediately.

```
task = "learn about uses of geospatial data"
task  # print the result
```

'learn about uses of geospatial data'

```
text <- "this also works"
print(text)
```

```
[1] "this also works"
```

- Arguments can optionally be named using =, and outputs of functions can be chained together.

```
x = rnorm(mean=0, sd=1, n=1000) + rnorm(mean=2, sd=1, n=1000)
head(x)  # print first 5 values
hist(x)  # make a histogram
```

1.19209914103509

0.334443797619174

4.56348110094114

1.18197851758601

0.324403505190461

-0.953382684526003

We'll see more examples of R below. The free W3Schools interactive tutorial is a great place to start if you want to learn more about R.

Opening data

Let's start with geospatial data in a simple format. The Neighborhood Deprivation Index (NDI) contains many useful sociodemographic covariates summarized at the census tract level.

```
# download geospatial data example
download.file('https://gis.cancer.gov/research/NeighDeprvIndex_USTracts.csv',
destfile='ndi.csv')

# inspect the file in our working directory (you should see "ndi.csv")
list.files(getwd())
```

'library'

'ndi.csv'

'rlib.tar.gz'

'sample_data'

Now that we've downloaded and unzipped the file, we can import it into an object in R. The read.csv() function used here will handle most kinds of geospatial data files. There are R packages available to handle other kinds of geospatial data - check out CRAN for more info.

```
ndi = read.csv('ndi.csv')

# verify import, notice that this is currently just a regular dataset
head(ndi)
```

A data.frame: 6 × 18

TractID

StCoFIPS

StAbbr

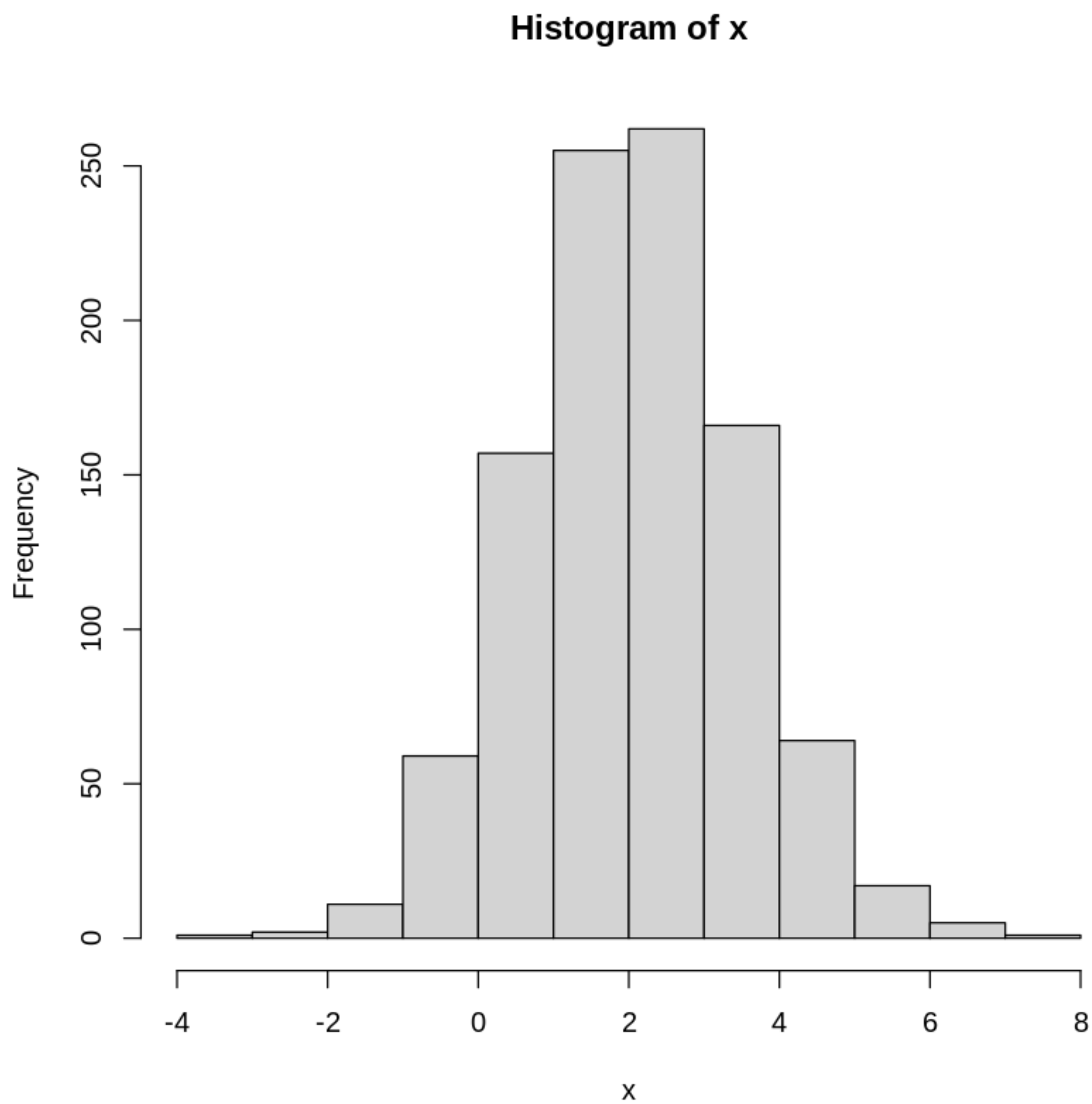

Figure 1: png

NDI

NDIQuint

MedHHInc

PctRecvIDR

PctPubAsst

MedHomeVal

PctMgmtBusSciArt

PctFemHeadKids

PctOwnerOcc

PctNoPhone

PctNComPlmb

PctEducHSPlus

PctEducBchPlus

PctFamBelowPov

PctUnempl

<dbl>

<int>

<chr>

<dbl>

<chr>

<int>

<dbl>

<dbl>

<int>

<dbl>

<dbl>

<dbl>

<dbl>

<dbl>

<dbl>

<dbl>

<dbl>

<dbl>

1

1001020100

1001

AL

-0.3082017

2-BelowAvg deprivation

67826

26.92307

12.068960

152500

38.47900

6.498673

75.2

2.7

0

90.6

37.8

12.0

4.6

2

1001020200

1001

AL

0.7938726

4-AboveAvg deprivation

41287

10.85568

24.137930

96100

30.51643

14.942520

61.9

1.1

0

82.0

16.2

18.3

3.4

3

1001020300

1001

AL

0.6497080

4-AboveAvg deprivation

46806

12.27521

12.900700

98900

27.86774

9.695074

66.4

0.7

0

86.3

18.1

10.0

4.7

4

1001020400

1001

AL

0.1747214

3-Average deprivation

55895

17.83876

5.660377

140800

28.98864

3.659233

79.6

1.6

1

90.0

26.7

1.5

6.1

5

1001020500

1001

AL

-0.5933945

2-BelowAvg deprivation

68143

20.17160

8.798283

187900

48.84060

3.505007

52.8

2.1

0

94.1

40.5

8.4

2.3

6

1001020600

1001

AL

0.6883914

4-AboveAvg deprivation

44549

17.84073

16.003060

93300

25.21994

8.039816

77.2

0.0

0

81.6

20.4

7.0

6.1

: A data.frame: 6 × 18

Making it spatial

How do we get the data to be "geospatial"? We'll need to add information about the geographic shapes of each of these rows (census tracts). Fortunately we can grab this information easily from the Census Bureau using the tigris package. We'll just look at one state to make things simple.

```
# first, cache (store) temporary results so we don't annoy the census bureau
# with many download requests if we need to run our code multiple times
options(tigris_use_cache = TRUE)

# next, get the data for maryland
md_tracts = tracts(year=2017, state='MD', cb=TRUE)

# inspect the top of the data frame, we use data.frame to show head()
# that we want to treat it as a data.frame (a tabular dataset)
head(data.frame(md_tracts))
```

```
|======================================================================| 100%
```

A data.frame: 6 × 10

STATEFP

COUNTYFP

TRACTCE

AFFGEOID

GEOID

NAME

LSAD

ALAND

AWATER

geometry

<chr>

<chr>

<chr>

<chr>

<chr>

<chr>

<chr>

<dbl>

<dbl>

<MULTIPOLYGON [°]>

1

24

003

702500

1400000US24003702500

24003702500

7025

CT

7347540

2848979

MULTIPOLYGON (((-76.55502 3...

2

24

003

706401

1400000US24003706401

24003706401

7064.01

CT

2216282

19353

MULTIPOLYGON (((-76.51287 3...

3

24

003

731003

1400000US24003731003

24003731003

7310.03

CT

3181054

1225330

MULTIPOLYGON (((-76.45633 3...

4

24

003

750101

1400000US24003750101

24003750101

7501.01

CT

1345541

0

MULTIPOLYGON (((-76.61357 3...

5

24

005

400600

1400000US24005400600

24005400600

4006

CT

968869

0

MULTIPOLYGON (((-76.72765 3…

6

24

005

400800

1400000US24005400800

24005400800

4008

CT

1472876

0

MULTIPOLYGON (((-76.74431 3…

: A data.frame: 6 × 10

Next, we'll join the census tract data to the svi dataset so we can see where things are. Notice the similarity between the 11-digit TractID in the ndi dataset and the 11-digit GEOID in the census_tracts dataset. Let's see if all the values in our maryland dataset are in the ndi dataset.

The only issue is that one of these variables is a `<chr>` and the other is a `<dbl>` (see the headers of the table above), so we'll convert the version in the tracts data to be a character variable.

```
# convert type to match
ndi$TractID = as.character(ndi$TractID)

# check types
```

```r
print(class(ndi$TractID))
print(class(md_tracts$GEOID))

# how many maryland tracts are in the id list
table(md_tracts$GEOID %in% ndi$TractID)  # looks good
```

```
[1] "character"
[1] "character"
```

```
TRUE
1396
```

Now, we can left-join the ndi data to the md_tracts. Left-join means that all records on the left (first) side form the basis for the dataset and the records on the right side are matched if possible (and dropped if not).

```r
# join dataset
md_ndi = left_join(
  md_tracts,
  ndi,
  by = join_by(GEOID==TractID)
)

# inspect difference in new dataset (27 vs 10 variables)
dim(md_tracts)
dim(md_ndi)
```

1396

10

1396

27

We're linked!

Inspecting data

Now that our dataset is set up, let's see what's there. The codebook has detailed descriptions of these variables. Notice that we still have some additional variables that were present in the census tract dataset.

```
# inspect variable names
names(md_ndi)
```

'STATEFP'

'COUNTYFP'

'TRACTCE'

'AFFGEOID'

'GEOID'

'NAME'

'LSAD'

'ALAND'

'AWATER'

'StCoFIPS'

'StAbbr'

'NDI'

'NDIQuint'

'MedHHInc'

'PctRecvIDR'

'PctPubAsst'

'MedHomeVal'

'PctMgmtBusSciArt'

'PctFemHeadKids'

'PctOwnerOcc'

'PctNoPhone'

'PctNComPlmb'

'PctEducHSPlus'

‘PctEducBchPlus’

‘PctFamBelowPov’

‘PctUnempl’

‘geometry’

The dataset is too large to look at all at once, so we’ll just inspect the first few rows to get a sense of the data. Notice that the geospatial data for each census tract is stored in the final variable called geometry (the data type for this column is MULTIPOLYGON) but only the first coordinate value is displayed in this view.

```
options(repr.matrix.max.cols = 100)  # override the default, and show all cols
head(data.frame(md_ndi))
```

A data.frame: 6 × 27

STATEFP

COUNTYFP

TRACTCE

AFFGEOID

GEOID

NAME

LSAD

ALAND

AWATER

StCoFIPS

StAbbr

NDI

NDIQuint

MedHHInc

PctRecvIDR

PctPubAsst

MedHomeVal

PctMgmtBusSciArt

PctFemHeadKids

PctOwnerOcc

PctNoPhone

PctNComPlmb

PctEducHSPlus

PctEducBchPlus

PctFamBelowPov

PctUnempl

geometry

<chr>

<chr>

<chr>

<chr>

<chr>

<chr>

<chr>

<dbl>

<dbl>

<int>

<chr>

<dbl>

<chr>

<int>

<dbl>

<dbl>

<int>

<dbl>

<dbl>

<dbl>

<dbl>

<dbl>

<dbl>

<dbl>

<dbl>

<dbl>

<MULTIPOLYGON [°]>

1

24

003

702500

1400000US24003702500

24003702500

7025

CT

7347540

2848979

24003

MD

-0.7913926

2-BelowAvg deprivation

88482

25.22608

19.228930

509400

43.63810

16.563540

50.8

2.3

0.0

85.4

36.1

11.8

6.9

MULTIPOLYGON (((-76.55502 3...

2

24

003

706401

1400000US24003706401

24003706401

7064.01

CT

2216282

19353

24003

MD

0.1526330

3-Average deprivation

69945

13.79000

12.472550

254200

28.06893

7.026789

48.8

0.0

0.0

70.1

27.9

8.0

4.3

MULTIPOLYGON (((-76.51287 3…

3

24

003

731003

1400000US24003731003

24003731003

7310.03

CT

3181054

1225330

24003

MD

-1.1298226

1-Least deprivation

91750

31.42857

3.376623

364100

46.26802

5.454545

96.2

1.0

0.0

96.4

45.4

0.6

2.9

MULTIPOLYGON (((-76.45633 3...

4

24

003

750101

1400000US24003750101

24003750101

7501.01

CT

1345541

0

24003

MD

0.2671614

3-Average deprivation

55492

25.07997

22.008950

198300

29.36378

11.772230

70.7

2.9

0.5

85.5

16.3

20.2

8.0

MULTIPOLYGON (((-76.61357 3...

5

24

005

400600

1400000US24005400600

24005400600

4006

CT

968869

0

24005

MD

-1.0638558

1-Least deprivation

85583

26.03806

9.861591

241300

51.23734

5.882352

66.3

0.9

0.0

95.3

53.2

1.5

3.7

MULTIPOLYGON (((-76.72765 3...

6

24

005

400800

1400000US24005400800

24005400800

4008

CT

1472876

0

24005

MD

-0.6472177

2-BelowAvg deprivation

88864

33.88157

12.061400

273600

44.31818

6.578947

66.1

3.6

1.4

87.8

30.4

1.3

6.6

MULTIPOLYGON (((-76.74431 3...

: A data.frame: 6 × 27

Geometry

What's actually "in" this mysterious `geometry` variable? Let's take a quick look at the first observation, using `$` to specify a variable and `[[1]]` to grab its first value.

```
md_ndi$geometry[[1]]
```

```
MULTIPOLYGON (((-76.55502 38.95145, -76.55338 38.95381, -76.55326 38.95398, -76.55156 38.9568
```

It appears to be a sequence of coordinate values. Let's explore a bit further and see if we can plot them. We'll use better tools to do this below, but it's nice to see that there's nothing "magic" going on here.

```
# extract one value
raw_data = unlist(md_ndi$geometry[[1]]) # force the object into a vector
n = length(raw_data)                    # find the length
x = raw_data[1:(n/2)]                   # the first half is the longitude (x)
y = raw_data[(n/2 + 1):n]               # the second half is the latitude (y)

# plot as lines connecting points
options(repr.plot.width = 8, repr.plot.height = 8)
plot(x,y, type='b')
```

Distributions

The md_ndi object that we created acts like a normal R dataset. For instance, we can look at the histogram of any of these variables using ggplot.

It's worth noting a few things about the syntax for ggplot() because we'll be using it below to create some maps:

- The main ggplot() call specifies the data we want to plot.
- geom_histogram() specifies the graph style.
- aes() specifies the aesthetics of the graph: how variables in the dataset are used (i.e. the x-axis is set to the NDI value).
- binwidth sets the graph to show small increments (size 0.1).
- labs() adds labels to elements of the graph.

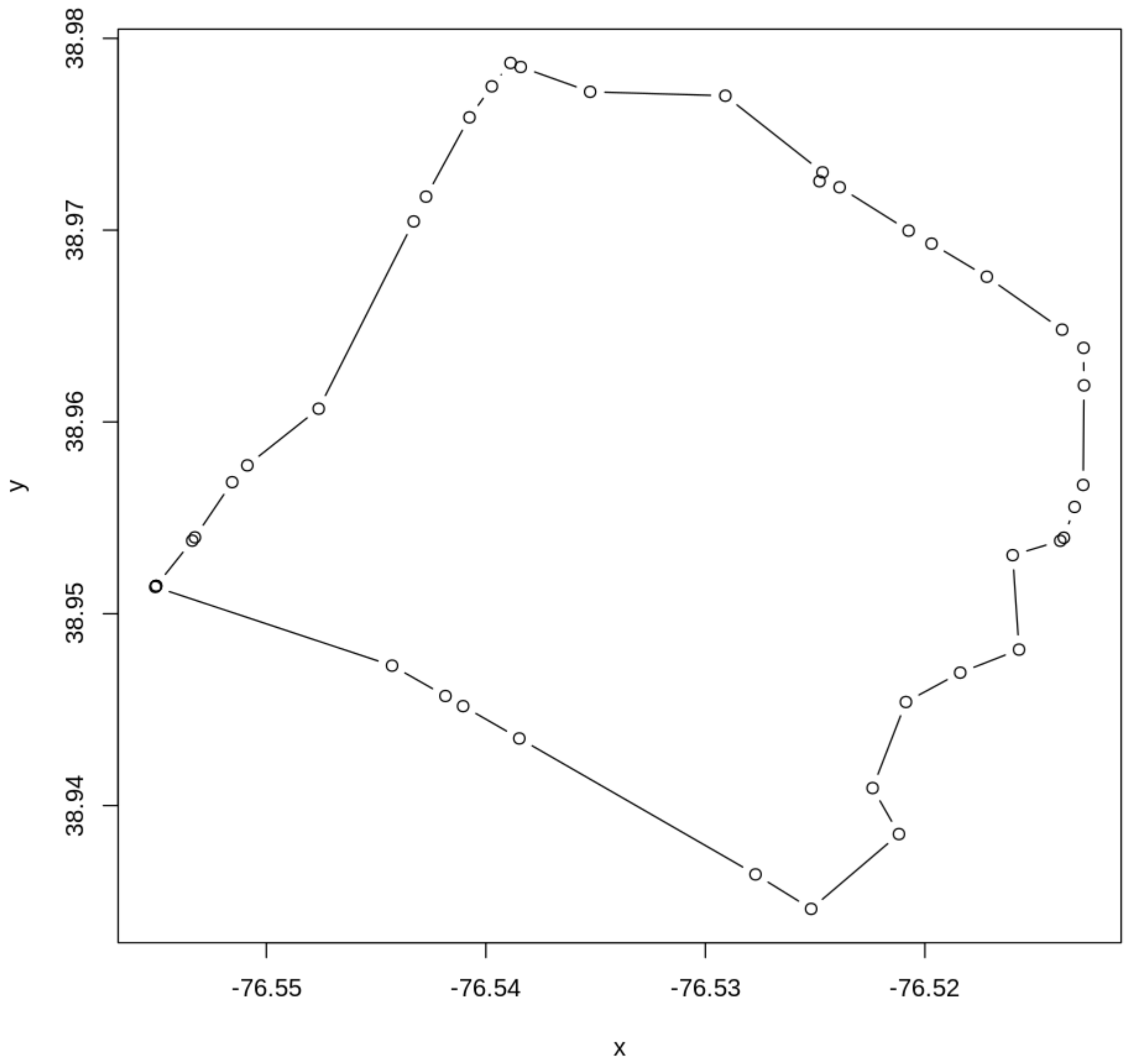

Figure 2: png

- Notice that we've broken the code up into multiple lines by giving incomplete lines ending in + that R combines.
- Also notice we get a warning message about "non-finite values" being removed - this is R warning us that there are missing values not being plotted (this is okay, but it's good that R tells us about it)

```
ggplot(md_ndi) +
  geom_histogram(aes(x=NDI), binwidth=0.1) +
  labs(title='Figure 1: Distribution of NDI', x='Neighborhood Deprivation Index')
```

```
Warning message:
"[1m[22mRemoved 93 rows containing non-finite outside the scale range (`stat_bin()`)."
```

Mapping

To create a map, we just need to change the graph options: * We first set options to increase the size of our plot window, to get a larger figure with more detail. If you right-click the figure and open it in a new tab you can see it at full size. * The type of graph geometry is now geom_sf, meaning spatial feature * Instead of assigning the variable to the x-axis coordinate, we use it to determine the fill color. * We set the line color for the plot to NA (missing) to prevent the lines from drawing on top of the fill colors, so we can see them better. * There aren't colors for areas without mail delivery or for which the variables could not be calculated due to low population (e.g. Western U.S. and Alaska), and these default to transparent.

```
options(repr.plot.width = 12, repr.plot.height = 8)
ggplot(md_ndi) +
  geom_sf(aes(fill=NDI), color=NA) +
  labs(title='Figure 2: Distribution of NDI in MD')
```

Deriving variables

Maps can be useful for understanding your data, but how do we integrate these variables into analysis models?

Typically, you'll start from residential addresses in your study that you geocode using a secure geocoder like DeGAUSS. The geocoding process takes a residential address and looks up the corresponding geographic coordinates (longitude=X and latitude=Y). There are privacy protection concerns at this step, so it is performed in a secure and protected computing environment.

You can see what a geocoder does by trying out the Census Geocoder.

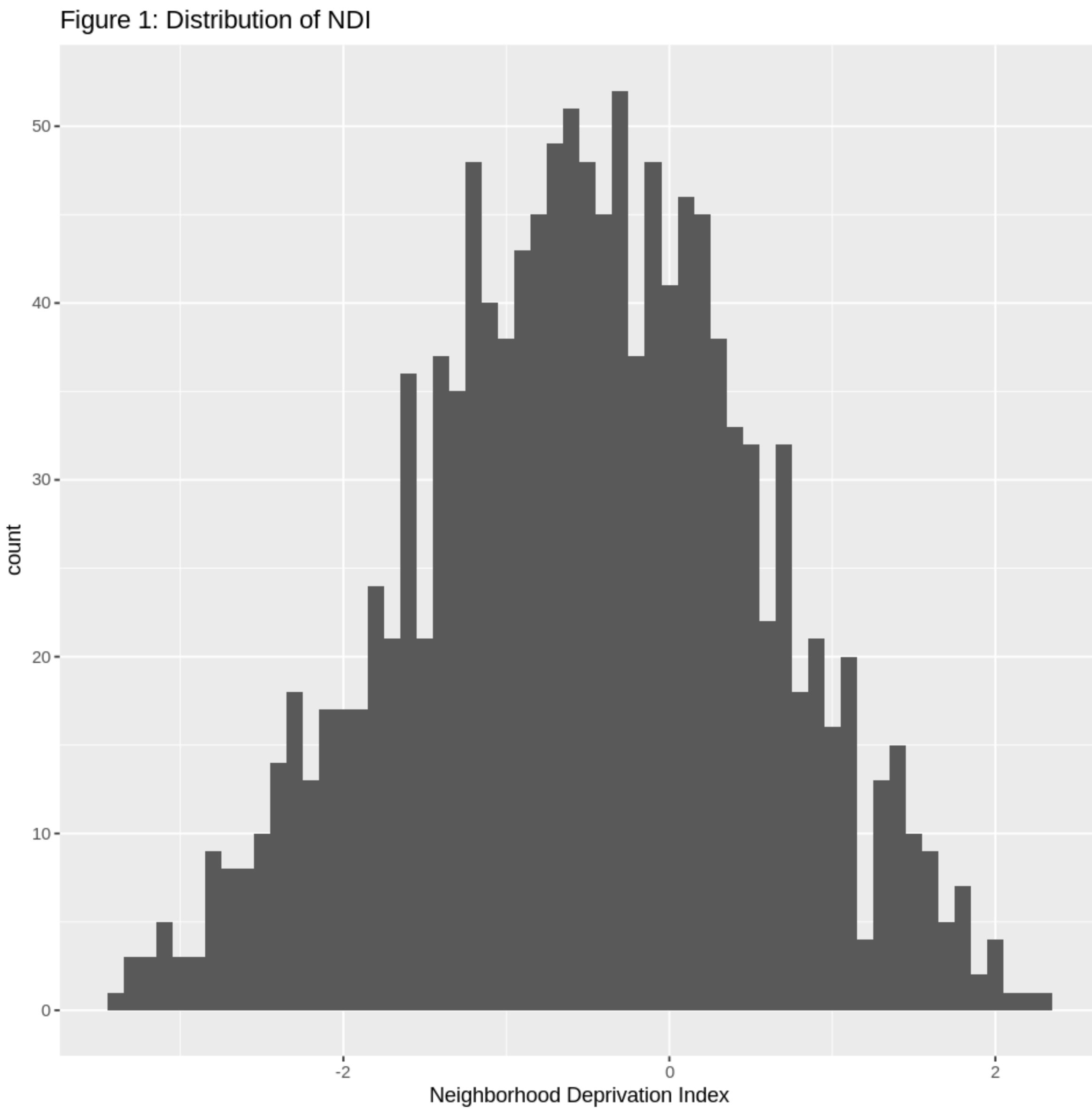
Figure 1: Distribution of NDI
count
50
40
30
20
10
0
-2
0
2
Neighborhood Deprivation Index

Figure 3: png

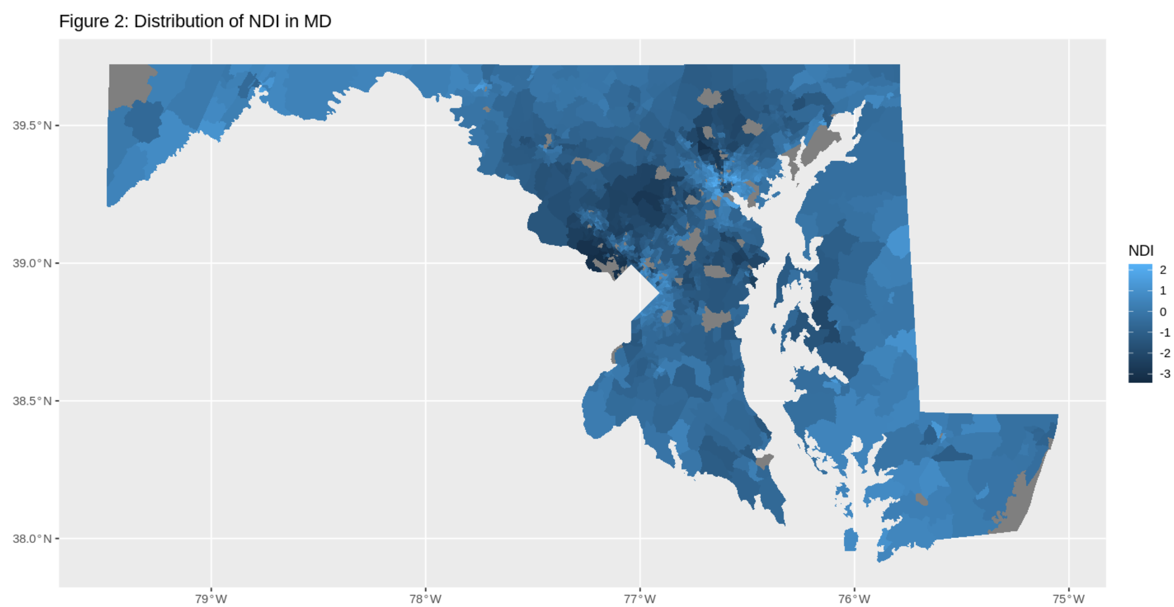
Figure 2: Distribution of NDI in MD
39.5°N
39.0°N
38.5°N
38.0°N
79°W
78°W
77°W
76°W
75°W
NDI
2
1
0
-1
-2
-3

Figure 4: png

Simulate a cohort

For this activity, we'll pretend that you've already passed the study addresses through a geocoder and have latitude and longitude coordinates for each participant. We'll simulate a basic cohort study dataset to mimic what you might start with before doing the geographic linkage.

```
# simulate a study area
study_area = st_union(st_geometry(md_tracts))
plot(study_area)
```

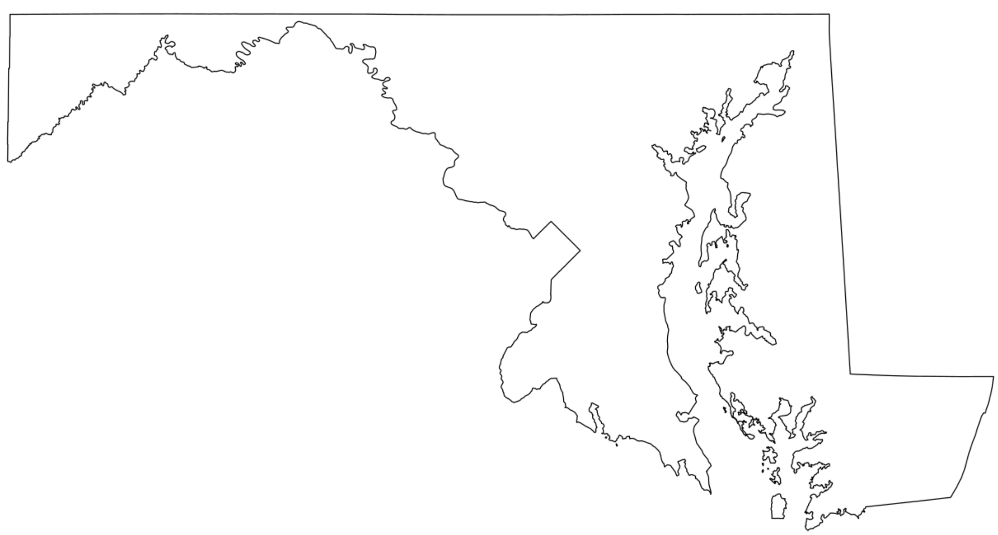

Figure 5: png

```
# simulate a study sample
N = 10000
study_sample = st_sample(study_area, size=N)

study_data = st_sf(
  data.frame(id=1:N),
```

```
  geometry=study_sample
)

# plot points on the same map
plot(study_area)
plot(study_sample, add=TRUE, pch='.')
```

```
Warning message in st_poly_sample(x, size = size, ..., type = type, by_polygon = by_polygon,
"coordinate ranges not computed along great circles; install package lwgeom to get rid of thi
Warning message in st_poly_sample(x, size = size, ..., type = type, by_polygon = by_polygon,
"coordinate ranges not computed along great circles; install package lwgeom to get rid of thi
```

Figure 6: png

Link data

Next, let's link the geospatial data to our cohort. Notice that we have an id number for each participant on the left, and the rest of the data corresponds to the information from the census tract within which that person is located.

```
linked = st_join(study_data, md_ndi)
head(data.frame(linked))
```

A data.frame: 6 × 28

id

STATEFP

COUNTYFP

TRACTCE

AFFGEOID

GEOID

NAME

LSAD

ALAND

AWATER

StCoFIPS

StAbbr

NDI

NDIQuint

MedHHInc

PctRecvIDR

PctPubAsst

MedHomeVal

PctMgmtBusSciArt

PctFemHeadKids

PctOwnerOcc

PctNoPhone

PctNComPlmb

PctEducHSPlus

PctEducBchPlus

PctFamBelowPov

PctUnempl

geometry

<int>

<chr>

<chr>

<chr>

<chr>

<chr>

<chr>

<chr>

<dbl>

<dbl>

<int>

<chr>

<dbl>

<chr>

<int>

<dbl>

<dbl>

<int>

<dbl>

<dbl>

<dbl>

<dbl>

<dbl>

<dbl>

<dbl>

<dbl>

<dbl>

<POINT [°]>

1

1

24

043

010200

1400000US24043010200

24043010200

102

CT

84272781

73363

24043

MD

-0.4580011

2-BelowAvg deprivation

76769

28.67602

9.517998

249500

39.84960

3.843807

86.8

2.6

0.4

88.5

30.5

3.6

6.8

POINT (-77.5788 39.69493)

2

2

24

003

701300

1400000US24003701300

24003701300

7013

CT

101097581

0

24003

MD

-1.4257690

1-Least deprivation

119844

35.81640

6.170052

509300

52.21526

3.912716

91.0

0.9

0.0

93.3

44.9

3.9

4.9

POINT (-76.66315 38.83531)

3

3

24

021

775302

1400000US24021775302

24021775302

7753.02

CT

41293126

2230283

24021

MD

-0.8795162

2-BelowAvg deprivation

96750

27.15719

7.357859

267500

46.91023

7.290969

82.1

0.5

0.0

92.4

42.5

5.1

6.6

POINT (-77.66202 39.34485)

4

4

24

043

011201

1400000US24043011201

24043011201

112.01

CT

16333161

0

24043

MD

-0.2893017

2-BelowAvg deprivation

59141

18.96690

14.474030

249500

36.58371

5.622814

48.4

0.0

0.0

90.7

34.7

3.4

8.1

POINT (-77.66543 39.61838)

5

5

24

033

800800

1400000US24033800800

24033800800

8008

CT

106692251

4690419

24033

MD

-1.0088142

1-Least deprivation

116919

34.05526

4.182225

416700

46.16204

0.746825

90.7

4.0

3.7

87.4

33.8

2.8

1.9

POINT (-76.7382 38.70385)

6

6

24

029

950100

1400000US24029950100

24029950100

9501

CT

170616697

5541555

24029

MD

-0.3668521

2-BelowAvg deprivation

61494

24.70198

9.072847

230700

41.00755

5.827814

74.8

4.0

0.0

87.2

32.8

5.1

6.3

POINT (-75.89985 39.25638)

: A data.frame: 6 × 28

Study Impacts

Simulate a trial

To see how this would affect a real study, let's give our cohort some additional study data and simulate a treatment and outcomes. This directed acyclic graph shows a hypothetical study setup:

Figure 7: DAG.svg

For this demonstration, we'll assume a simple causal model where our exposure in influenced by ndi and the outcome is influenced by ndi and the exposure. * This is similar to how in an RCT, treatment assignment is randomized but actual treatment can't be effectively randomized because we can't directly control adherence (the intent-to-treat and as-treated effects differ). * We'll model treatment effect as having a fixed value, in reality, the treatment effect different for each person, and we're aiming to estimate various versions of the "mean" effect in different populations (i.e. among the cohort, the treated, standardized to a specific population, etc.).

Here's the simpler and more general form we'll simulate:

```
# ## Set simulation parameters
# We'll start out with a bit of an "extreme" effect of the confound for illustration
# you can try changing the parameters in this section to see how it affects
# the bias of the crude model.
base_exposure_odds = 0.2      # base odds of exposure
base_outcome_odds = 0.2       # base odds of outcome
exposure_or = 2               # odds ratio: exposure -> outcome
ndi_exposure_or = 2           # odds ratio: ndi -> expoxure
ndi_or = 3                    # odds ration: ndi -> outcome


# create simulated data
cohort = data.frame(
  id=1:N,
```

```
  ndi = linked$NDI) %>%

  # drop observations without an NDI value
  filter(!is.na(ndi)) %>%

  # add other data
  mutate(

    # transform ndi into a binary variable for simplicity
    # ndi=1 means that the ndi is worse, 0 is better
    ndi = ndi > mean(ndi),

    # exposure assignment is based on ndi
    # (twice as likely for high NDI)
    exposure_odds = exp(
      log(base_exposure_odds) +
      log(ndi_exposure_or)*ndi),
    exposure_probability = exposure_odds / (1+exposure_odds),
    exposure = rbinom(p=exposure_probability, size=1, n=n()),

    # outcome is based on ndi and exposure
    outcome_odds = exp(
      log(base_outcome_odds) +
      log(exposure_or)*exposure +
      log(ndi_or)*ndi),
    outcome_probability = outcome_odds / (1+outcome_odds),
    outcome = rbinom(p=outcome_probability, size=1, n=n())

  )

# inspect the data structure
head(cohort, n=15)
```

Table 2: A data.frame: 15 × 8

	id	ndi	exposure_odds	exposure_probability	exposure	outcome_odds	outcome_probability	outcome
	<int>	<lgl>	<dbl>	<dbl>	<int>	<dbl>	<dbl>	<int>
1	1	TRUE	0.4	0.2857143	0	0.6	0.3750000	1
2	2	FALSE	0.2	0.1666667	0	0.2	0.1666667	0
3	3	FALSE	0.2	0.1666667	0	0.2	0.1666667	0
4	4	TRUE	0.4	0.2857143	1	1.2	0.5454545	0

	id	ndi	exposure_odds	exposure_probability	exposure	outcome_odds	outcome_probability	outcome
5	5	FALSE	0.2	0.1666667	0	0.2	0.1666667	0
6	6	TRUE	0.4	0.2857143	0	0.6	0.3750000	1
7	7	TRUE	0.4	0.2857143	0	0.6	0.3750000	0
8	8	TRUE	0.4	0.2857143	0	0.6	0.3750000	0
9	9	TRUE	0.4	0.2857143	0	0.6	0.3750000	0
10	10	TRUE	0.4	0.2857143	0	0.6	0.3750000	1
11	11	FALSE	0.2	0.1666667	1	0.4	0.2857143	0
12	12	FALSE	0.2	0.1666667	0	0.2	0.1666667	0
13	13	FALSE	0.2	0.1666667	0	0.2	0.1666667	0
14	14	TRUE	0.4	0.2857143	1	1.2	0.5454545	1
15	15	FALSE	0.2	0.1666667	0	0.2	0.1666667	0

Crude model

Let's take a look at our data and some crude models with no adjustment for spatial information. Notice that the 95% confidence interval is not centereed on the actual effect odds ratio.

```
crude_model = glm(outcome==1 ~ exposure, data=cohort, family=binomial('logit'))
summary(crude_model)

cat('\nTarget Odds Ratio:',exposure_or,'\n')
cat('\n\nOdds Ratios:\n')
exp(coef(crude_model))
exp(confint(crude_model))
```

```
Call:
glm(formula = outcome == 1 ~ exposure, family = binomial("logit"),
    data = cohort)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.01211    0.02646  -38.25   <2e-16 ***
exposure     0.84820    0.05040   16.83   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 11769  on 9492  degrees of freedom
```

```
Residual deviance: 11490  on 9491  degrees of freedom
AIC: 11494

Number of Fisher Scoring iterations: 4




Target Odds Ratio: 2


Odds Ratios:
```

(Intercept)

0.363449691992813

exposure

2.33544720567366

```
Waiting for profiling to be done...
```

Table 3: A matrix: 2 × 2 of type dbl

	2.5 %	97.5 %
(Intercept)	0.345006	0.3827188
exposure	2.115742	2.5779048

Adjusted model

What happens if we adjust for our spatial covariate, ndi? We can see that the estimate is closer to the true value set in the simulation.

```
adjusted_model = glm(outcome==1 ~ exposure + ndi, data=cohort, family=binomial('logit'))
summary(adjusted_model)

cat('\nTarget Odds Ratio:',exposure_or,'\n')
cat('\n\nOdds Ratios:\n')
exp(coef(adjusted_model))
exp(confint(adjusted_model))
```

```
Call:
glm(formula = outcome == 1 ~ exposure + ndi, family = binomial("logit"),
    data = cohort)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.68956    0.04238  -39.87   <2e-16 ***
exposure     0.70395    0.05222   13.48   <2e-16 ***
ndiTRUE      1.16383    0.04974   23.40   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 11769  on 9492  degrees of freedom
Residual deviance: 10898  on 9490  degrees of freedom
AIC: 10904

Number of Fisher Scoring iterations: 4
```

```
Target Odds Ratio: 2
```

```
Odds Ratios:
```

(Intercept)

0.18460001453816

exposure

2.0217225179843

ndiTRUE

3.20216368875125

```
Waiting for profiling to be done...
```

Table 4: A matrix: 3×2 of type dbl

	2.5 %	97.5 %
(Intercept)	0.1697653	0.2004487
exposure	1.8249467	2.2395504
ndiTRUE	2.9058970	3.5315380

Matching

An alternate approach to covariate adjusment is to match participants with similar geospatial values when making comparisons. This is particularly effective when you want to match on multiple geospatial variables, because if you can compare neighbors you know what all geospatial elements are by definition similar. This means that the approach also adjusts for for ***unmeasured*** geospatial covariates, an idea that is leveraged in many geostatistical methods.

```
# link the geospatial information ("linked") to our study dataset by ID
cohort_spatial = cohort %>%
  left_join(linked, by='id')

# see participant sets living in the same areas
groups = cohort_spatial %>%
  select(id, GEOID) %>%
  group_by(GEOID) %>%
  summarize(ids = paste(id, collapse=', '), n=n())

head(groups, n=15)

# look at the number of participants in each geographical area
# note that we link back to the polygon-based census tract file
groups_by_map = md_tracts %>%
  left_join(groups, by='GEOID')


options(repr.plot.width = 12, repr.plot.height = 8)
ggplot(groups_by_map) +
  geom_sf(aes(fill=n)) +
  labs(title="Figure 3: Number of Participants per Census Tract")
```

Table 5: A tibble: 15 × 3

GEOID	ids	n
<chr>	<chr>	<int>
[illegible]	47, 112, 121, 135, 141, 192, 399, 435, 460, 495, 524, 530, 545, 560, 729, 742, 826, 1083, 1099, 1136, 1228, 1321, 1435, 1452, 1509, 1543, 1550, 1562, 1715, 1784, 1828, 1862, 1909, 1999, 2039, 2076, 2084, 2086, 2095, 2139, 2227, 2279, 2450, 2515, 2537, 2616, 2623, 2666, 2709, 2740, 2870, 2917, 2947, 2949, 3006, 3079, 3097, 3315, 3351, 3358, 3542, 3597, 3634, 3771, 3795, 3850, 3939, 3969, 3979, 3988, 3994, 4061, 4079, 4080, 4115, 4288, 4318, 4364, 4420, 4421, 4450, 4613, 4707, 4780, 4799, 4834, 4846, 4991, 5001, 5130, 5194, 5217, 5235, 5472, 5505, 5507, 5525, 5721, 5745, 5772, 5779, 5868, 5968, 6011, 6151, 6201, 6227, 6232, 6262, 6301, 6365, 6462, 6570, 6601, 6733, 6734, 6788, 6915, 6922, 6938, 6969, 7009, 7017, 7025, 7109, 7162, 7238, 7295, 7350, 7420, 7488, 7567, 7571, 7640, 7655, 7682, 7761, 7793, 7836, 7854, 7892, 8066, 8068, 8165, 8234, 8280, 8333, 8354, 8576, 8624, 8696, 8709, 8738, 8763, 8840, 9028, 9040, 9091, 9092, 9275, 9303, 9372, 9393, 9396, 9445, 9463, 9505, 9512, 9564, 9666, 9745, 9843, 9943, 9975, 9980	175
[illegible]	[illegible], 690, 707, 784, 886, 912, 962, 995, 1386, 1572, 1579, 1644, 1881, 1928, 2026, 2255, 2627, 2915, 3671, 3687, 3727, 3809, 4136, 4835, 5196, 5335, 5679, 5960, 6025, 6970, 7065, 7080, 7108, 7329, 7717, 7889, 8137, 8194, 8403, 9356, 9459	42
[illegible]	[illegible], 6410, 9866	4
[illegible]	[illegible], 5106, 7848	4
[illegible]	[illegible]	2
[illegible]	[illegible]	2
[illegible]	[illegible]	2
[illegible]	[illegible], 4934	3
[illegible]	[illegible]	1
[illegible]	[illegible], 4574, 6754	4
[illegible]	[illegible], 5486, 6505, 7047, 8760	6
[illegible]	[illegible], 5580, 6346, 6566, 7005	6
[illegible]	[illegible], 5508, 5600, 6419, 7064, 7987, 8693	8
[illegible]	[illegible], 2024, 2057, 2404, 2885, 3372, 4681, 7426, 8536	10
[illegible]	[illegible], 1682, 1960, 2692, 5870, 6274, 7039, 7824, 9507	10

For this illustration we'll show a generalization of matching: stratification. Instead of splitting the cohort into many 2-person groups, we'll just use each geographic area as a group and run the comparison across groups ('strata').

```
cohort_with_strata = cohort %>%
  # join the geospatial identifiers
  left_join(linked %>% select(GEOID, id), by='id') %>%
```

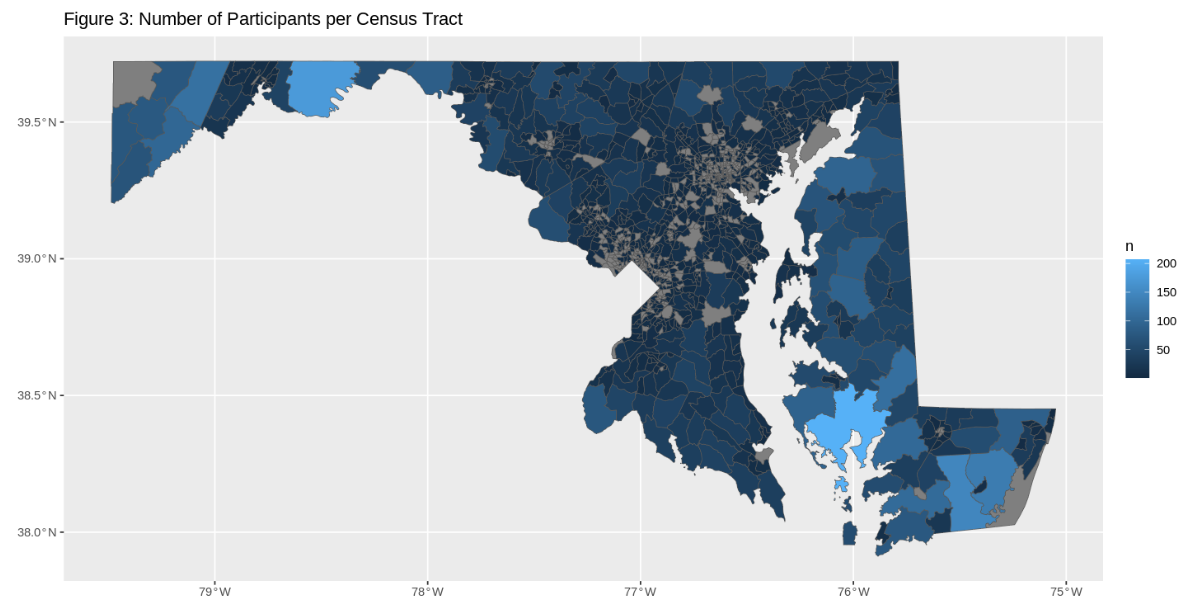
Figure 3: Number of Participants per Census Tract
39.5°N
39.0°N
38.5°N
38.0°N
79°W
78°W
77°W
76°W
75°W
n
200
150
100
50

Figure 8: png

```r
  # join info about groups
  left_join(groups_by_map %>% select(GEOID, n), by='GEOID') %>%
  # filter to areas with at least 5 participants to simplify model fit
  filter(n>=5)

# fit the stratified model using glm. Realistically, you would want to use
# a specialized fixed-effects estimation R package like lme4
stratified_model = glm(outcome==1 ~ exposure + factor(GEOID),
  data=cohort_with_strata,
  family=binomial('logit'))

summary(stratified_model)

# This shows all of the effects of the individual census tracts.
# Notice that the estimate for "exposure" is still pretty accurate,
# even though we aren't explicitly modeling ndi in the model but instead are
# stratifying by the geographic identifier
cat('\nTarget Odds Ratio:',exposure_or,'\n')
cat('\n\nOdds Ratios:\n')
exp(coef(stratified_model))
```

```
Call:
glm(formula = outcome == 1 ~ exposure + factor(GEOID), family = binomial("logit"),
    data = cohort_with_strata)

Coefficients:
                           Estimate Std. Error z value Pr(>|z|)
(Intercept)              -3.833e-01  1.544e-01  -2.483 0.013035 *
exposure                  7.143e-01  5.752e-02  12.418  < 2e-16 ***
factor(GEOID)24001000200 -4.907e-02  3.501e-01  -0.140 0.888539
factor(GEOID)24001001300  9.707e-01  8.851e-01   1.097 0.272771
factor(GEOID)24001001401  9.707e-01  8.851e-01   1.097 0.272771
factor(GEOID)24001001502  4.613e-01  7.570e-01   0.609 0.542317
factor(GEOID)24001001503 -7.735e-01  7.162e-01  -1.080 0.280111
factor(GEOID)24001001600 -3.922e-01  6.734e-01  -0.582 0.560304
factor(GEOID)24001001700  3.256e-01  6.001e-01   0.543 0.587392
factor(GEOID)24001001900 -6.473e-01  4.836e-01  -1.339 0.180724
factor(GEOID)24001002000 -4.331e-01  5.065e-01  -0.855 0.392528
factor(GEOID)24001002100 -5.075e-01  4.108e-01  -1.235 0.216759
factor(GEOID)24001002200 -3.175e-01  5.139e-01  -0.618 0.536744
factor(GEOID)24003701200 -1.618e+01  8.484e+02  -0.019 0.984781
factor(GEOID)24003701300 -1.180e+00  4.829e-01  -2.444 0.014537 *
```

factor(GEOID)24003701400	-2.831e+00	1.036e+00	-2.733	0.006274	**
factor(GEOID)24003702100	-2.158e+00	1.065e+00	-2.025	0.042838	*
factor(GEOID)24003702402	-1.760e+00	1.100e+00	-1.600	0.109573	
factor(GEOID)24003702500	-1.705e-01	9.351e-01	-0.182	0.855334	
factor(GEOID)24003702702	-8.208e-02	6.976e-01	-0.118	0.906338	
factor(GEOID)24003707001	-1.128e+00	6.835e-01	-1.650	0.098967	.
factor(GEOID)24003708001	-2.166e+00	1.060e+00	-2.044	0.040962	*
factor(GEOID)24003730100	-2.073e+00	1.071e+00	-1.935	0.052992	.
factor(GEOID)24003730902	-3.156e-01	7.549e-01	-0.418	0.675880	
factor(GEOID)24003731202	-6.214e-01	7.137e-01	-0.871	0.383916	
factor(GEOID)24003731203	-1.632e+01	9.675e+02	-0.017	0.986543	
factor(GEOID)24003731303	-3.099e-01	7.238e-01	-0.428	0.668562	
factor(GEOID)24003731306	-1.644e+01	7.846e+02	-0.021	0.983279	
factor(GEOID)24003740500	-1.634e+01	1.058e+03	-0.015	0.987674	
factor(GEOID)24003740702	-2.314e+00	1.052e+00	-2.199	0.027904	*
factor(GEOID)24003740800	-1.618e+01	9.069e+02	-0.018	0.985764	
factor(GEOID)24003740900	-1.618e+01	9.796e+02	-0.017	0.986820	
factor(GEOID)24003751200	-2.594e-01	5.968e-01	-0.435	0.663782	
factor(GEOID)24003751500	-1.003e+00	1.129e+00	-0.889	0.374173	
factor(GEOID)24003751600	2.075e-01	6.040e-01	0.344	0.731147	
factor(GEOID)24003751700	-1.486e+00	8.042e-01	-1.847	0.064688	.
factor(GEOID)24005402201	-1.876e+00	1.088e+00	-1.725	0.084549	.
factor(GEOID)24005402202	-1.618e+01	9.069e+02	-0.018	0.985764	
factor(GEOID)24005403803	-1.226e+00	1.106e+00	-1.108	0.267704	
factor(GEOID)24005404101	-1.630e+01	8.970e+02	-0.018	0.985502	
factor(GEOID)24005404402	-1.003e+00	1.129e+00	-0.889	0.374173	
factor(GEOID)24005404600	-1.764e+00	6.330e-01	-2.787	0.005328	**
factor(GEOID)24005404800	-1.992e+00	1.070e+00	-1.861	0.062780	.
factor(GEOID)24005404900	-1.093e+00	6.749e-01	-1.620	0.105258	
factor(GEOID)24005405000	-1.198e+00	3.864e-01	-3.100	0.001933	**
factor(GEOID)24005406000	-1.845e+00	5.547e-01	-3.326	0.000881	***
factor(GEOID)24005407002	-1.587e+00	5.085e-01	-3.121	0.001802	**
factor(GEOID)24005408100	-5.330e-01	8.508e-01	-0.626	0.530997	
factor(GEOID)24005408200	-1.027e+00	5.361e-01	-1.915	0.055490	.
factor(GEOID)24005408900	-1.170e+00	1.138e+00	-1.029	0.303635	
factor(GEOID)24005410100	-9.060e-01	4.160e-01	-2.178	0.029427	*
factor(GEOID)24005410200	-1.534e+00	7.831e-01	-1.959	0.050135	.
factor(GEOID)24005411101	-1.644e+01	9.609e+02	-0.017	0.986347	
factor(GEOID)24005411102	-1.408e+00	1.091e+00	-1.291	0.196748	
factor(GEOID)24005411202	-1.369e+00	7.947e-01	-1.723	0.084853	.
factor(GEOID)24005411302	-1.369e+00	1.113e+00	-1.230	0.218675	
factor(GEOID)24005451701	-8.655e-01	8.616e-01	-1.005	0.315135	
factor(GEOID)24005451801	5.166e-01	9.386e-01	0.550	0.582003	

```
factor(GEOID)24005451900  9.707e-01  8.851e-01   1.097 0.272771
factor(GEOID)24005490100 -1.618e+01  9.796e+02  -0.017 0.986820
factor(GEOID)24009860101 -2.236e+00  1.061e+00  -2.108 0.035032 *
factor(GEOID)24009860102 -1.651e+00  1.099e+00  -1.502 0.133140
factor(GEOID)24009860200 -1.249e+00  5.852e-01  -2.135 0.032752 *
factor(GEOID)24009860300 -1.370e-01  6.123e-01  -0.224 0.822903
factor(GEOID)24009860401 -1.348e+00  8.036e-01  -1.678 0.093438 .
factor(GEOID)24009860402  1.791e+00  1.113e+00   1.609 0.107570
factor(GEOID)24009860501  3.339e-01  5.583e-01   0.598 0.549780
factor(GEOID)24009860502 -1.618e+01  9.796e+02  -0.017 0.986820
factor(GEOID)24009860600 -2.314e+00  1.052e+00  -2.199 0.027904 *
factor(GEOID)24009860701 -9.652e-01  6.040e-01  -1.598 0.110013
factor(GEOID)24009860702 -7.432e-01  7.002e-01  -1.061 0.288471
factor(GEOID)24009860703 -8.434e-01  4.767e-01  -1.769 0.076893 .
factor(GEOID)24009860801 -1.217e+00  4.749e-01  -2.563 0.010363 *
factor(GEOID)24009860802 -4.640e-01  7.071e-01  -0.656 0.511699
factor(GEOID)24009860900 -8.330e-01  5.036e-01  -1.654 0.098122 .
factor(GEOID)24009861001 -1.003e+00  8.055e-01  -1.245 0.213060
factor(GEOID)24011955000 -7.140e-01  4.082e-01  -1.749 0.080282 .
factor(GEOID)24011955100  3.356e-01  3.374e-01   0.995 0.319911
factor(GEOID)24011955201  2.316e-01  3.850e-01   0.602 0.547441
factor(GEOID)24011955202 -6.669e-01  3.836e-01  -1.738 0.082134 .
factor(GEOID)24011955301  1.291e-01  3.848e-01   0.335 0.737312
factor(GEOID)24011955400 -7.064e-01  3.699e-01  -1.910 0.056188 .
factor(GEOID)24011955500 -1.539e-01  3.435e-01  -0.448 0.654092
factor(GEOID)24011955600 -2.204e-01  3.693e-01  -0.597 0.550635
factor(GEOID)24013501001 -7.581e-01  8.605e-01  -0.881 0.378326
factor(GEOID)24013501002 -6.076e-02  4.155e-01  -0.146 0.883758
factor(GEOID)24013502000 -2.174e+00  7.534e-01  -2.885 0.003909 **
factor(GEOID)24013503000 -3.325e-01  3.349e-01  -0.993 0.320838
factor(GEOID)24013504100 -1.549e+00  7.771e-01  -1.993 0.046287 *
factor(GEOID)24013504201 -9.623e-01  8.209e-01  -1.172 0.241114
factor(GEOID)24013504202 -1.122e+00  5.987e-01  -1.874 0.060931 .
factor(GEOID)24013505102 -1.632e+01  9.675e+02  -0.017 0.986543
factor(GEOID)24013505206 -1.618e+01  9.069e+02  -0.018 0.985764
factor(GEOID)24013506101 -7.973e-01  5.416e-01  -1.472 0.140960
factor(GEOID)24013506102 -3.958e-01  7.281e-01  -0.544 0.586749
factor(GEOID)24013506200 -2.750e+00  1.035e+00  -2.658 0.007865 **
factor(GEOID)24013507500 -2.215e-01  7.514e-01  -0.295 0.768127
factor(GEOID)24013507601 -5.653e-01  6.395e-01  -0.884 0.376647
factor(GEOID)24013507602  4.553e-01  6.949e-01   0.655 0.512332
factor(GEOID)24013507702 -3.141e-01  6.512e-01  -0.482 0.629603
factor(GEOID)24013507802 -9.169e-01  8.393e-01  -1.092 0.274627
```

factor(GEOID)24013508101	-2.236e+00	1.061e+00	-2.108	0.035032	*
factor(GEOID)24013508200	-6.091e-01	5.095e-01	-1.196	0.231868	
factor(GEOID)24013509001	-8.155e-01	5.481e-01	-1.488	0.136756	
factor(GEOID)24013509002	-2.449e-01	6.720e-01	-0.364	0.715563	
factor(GEOID)24013510000	-4.465e-01	5.367e-01	-0.832	0.405399	
factor(GEOID)24013511000	-1.685e+00	7.701e-01	-2.189	0.028624	*
factor(GEOID)24013513002	-1.093e+00	6.749e-01	-1.620	0.105258	
factor(GEOID)24013514201	-1.027e+00	5.863e-01	-1.752	0.079767	.
factor(GEOID)24015030100	-2.750e-01	2.947e-01	-0.933	0.350812	
factor(GEOID)24015030200	-7.100e-01	3.798e-01	-1.869	0.061578	.
factor(GEOID)24015030501	-3.169e-01	5.980e-01	-0.530	0.596128	
factor(GEOID)24015030503	-8.004e-01	4.476e-01	-1.788	0.073700	.
factor(GEOID)24015030505	-4.386e-01	8.872e-01	-0.494	0.621081	
factor(GEOID)24015030601	-4.386e-01	4.632e-01	-0.947	0.343709	
factor(GEOID)24015030602	-1.226e+00	1.106e+00	-1.108	0.267704	
factor(GEOID)24015030700	-5.757e-01	5.001e-01	-1.151	0.249704	
factor(GEOID)24015030903	8.029e-02	6.924e-01	0.116	0.907677	
factor(GEOID)24015030904	-1.262e+00	4.252e-01	-2.969	0.002986	**
factor(GEOID)24015030905	-1.417e-01	5.342e-01	-0.265	0.790840	
factor(GEOID)24015030906	-2.450e-01	4.792e-01	-0.511	0.609107	
factor(GEOID)24015031201	-2.820e-01	4.880e-01	-0.578	0.563369	
factor(GEOID)24015031202	-3.156e-01	7.549e-01	-0.418	0.675880	
factor(GEOID)24015031301	-6.884e-01	6.400e-01	-1.076	0.282121	
factor(GEOID)24015031302	-3.738e-01	5.693e-01	-0.657	0.511419	
factor(GEOID)24015031400	1.912e-01	4.583e-01	0.417	0.676516	
factor(GEOID)24017850101	-9.611e-02	4.842e-01	-0.198	0.842676	
factor(GEOID)24017850102	-1.636e-01	5.059e-01	-0.323	0.746468	
factor(GEOID)24017850201	-1.610e-01	5.859e-01	-0.275	0.783538	
factor(GEOID)24017850300	-3.827e-01	6.531e-01	-0.586	0.557864	
factor(GEOID)24017850400	-1.188e-01	2.831e-01	-0.420	0.674749	
factor(GEOID)24017850500	-1.387e+00	4.407e-01	-3.147	0.001651	**
factor(GEOID)24017850600	-1.812e+00	6.403e-01	-2.830	0.004658	**
factor(GEOID)24017850712	-8.175e-01	8.361e-01	-0.978	0.328196	
factor(GEOID)24017850713	-1.644e+01	9.609e+02	-0.017	0.986347	
factor(GEOID)24017850802	-6.911e-02	4.528e-01	-0.153	0.878680	
factor(GEOID)24017850904	-1.779e+00	7.763e-01	-2.291	0.021952	*
factor(GEOID)24017851001	-1.369e+00	1.113e+00	-1.230	0.218675	
factor(GEOID)24017851002	-1.623e+01	5.790e+02	-0.028	0.977636	
factor(GEOID)24017851100	-1.775e-01	4.046e-01	-0.439	0.660852	
factor(GEOID)24017851200	4.607e-02	3.675e-01	0.125	0.900248	
factor(GEOID)24017851301	-1.383e+00	4.457e-01	-3.104	0.001911	**
factor(GEOID)24017851302	-7.072e-01	5.061e-01	-1.397	0.162322	
factor(GEOID)24017851400	-1.329e+00	4.437e-01	-2.996	0.002732	**

```
factor(GEOID)24017851500 -1.634e+01  1.058e+03  -0.015 0.987674
factor(GEOID)24019970100 -5.419e-02  2.468e-01  -0.220 0.826188
factor(GEOID)24019970200  3.029e-02  3.042e-01   0.100 0.920681
factor(GEOID)24019970300 -9.833e-02  3.303e-01  -0.298 0.765931
factor(GEOID)24019970600 -3.156e-01  7.549e-01  -0.418 0.675880
factor(GEOID)24019970702  2.274e-01  3.152e-01   0.722 0.470509
factor(GEOID)24019970804  3.628e-01  2.618e-01   1.386 0.165799
factor(GEOID)24019970900 -1.380e-01  2.097e-01  -0.658 0.510414
factor(GEOID)24021740200 -7.581e-01  8.605e-01  -0.881 0.378326
factor(GEOID)24021751003 -1.326e+00  8.136e-01  -1.629 0.103272
factor(GEOID)24021751004 -1.226e+00  1.106e+00  -1.108 0.267704
factor(GEOID)24021751301 -1.797e+00  7.683e-01  -2.339 0.019360 *
factor(GEOID)24021751302 -4.680e-01  3.720e-01  -1.258 0.208306
factor(GEOID)24021751600  3.758e-01  3.620e-01   1.038 0.299240
factor(GEOID)24021751701 -1.170e+00  1.138e+00  -1.029 0.303635
factor(GEOID)24021751702 -2.292e+00  6.202e-01  -3.697 0.000219 ***
factor(GEOID)24021751801 -5.902e-01  6.138e-01  -0.962 0.336259
factor(GEOID)24021751802 -2.121e+00  7.610e-01  -2.787 0.005318 **
factor(GEOID)24021751902 -1.170e+00  1.138e+00  -1.029 0.303635
factor(GEOID)24021751904 -2.185e-01  7.908e-01  -0.276 0.782316
factor(GEOID)24021752001 -2.218e-02  9.258e-01  -0.024 0.980890
factor(GEOID)24021752102 -1.769e+00  7.806e-01  -2.266 0.023427 *
factor(GEOID)24021752201 -1.866e+00  6.326e-01  -2.949 0.003189 **
factor(GEOID)24021752204 -1.618e+01  9.796e+02  -0.017 0.986820
factor(GEOID)24021752302 -8.888e-01  4.667e-01  -1.904 0.056849 .
factor(GEOID)24021752303 -8.292e-03  5.657e-01  -0.015 0.988304
factor(GEOID)24021752501 -6.567e-01  4.528e-01  -1.450 0.147009
factor(GEOID)24021752502 -1.618e+01  6.413e+02  -0.025 0.979868
factor(GEOID)24021752601 -1.634e+01  1.058e+03  -0.015 0.987674
factor(GEOID)24021752602 -1.128e+00  6.835e-01  -1.650 0.098967 .
factor(GEOID)24021752801 -5.801e-01  4.824e-01  -1.202 0.229186
factor(GEOID)24021752802 -1.369e+00  6.549e-01  -2.091 0.036529 *
factor(GEOID)24021752900  3.535e-02  3.163e-01   0.112 0.911023
factor(GEOID)24021766800  1.076e+00  8.797e-01   1.224 0.221076
factor(GEOID)24021767500 -2.028e-01  3.467e-01  -0.585 0.558716
factor(GEOID)24021767600 -1.190e+00  4.274e-01  -2.785 0.005351 **
factor(GEOID)24021770700 -8.586e-01  5.375e-01  -1.597 0.110163
factor(GEOID)24021775302 -8.330e-01  5.036e-01  -1.654 0.098122 .
factor(GEOID)24023000200 -2.575e-01  2.850e-01  -0.903 0.366277
factor(GEOID)24023000300 -4.135e-01  2.500e-01  -1.654 0.098131 .
factor(GEOID)24023000400 -3.827e-01  2.616e-01  -1.463 0.143475
factor(GEOID)24023000500 -1.380e+00  3.424e-01  -4.030 5.57e-05 ***
factor(GEOID)24023000600 -2.273e-01  2.858e-01  -0.795 0.426392
```

```
factor(GEOID)24023000700  2.627e-02  2.891e-01   0.091 0.927598
factor(GEOID)24025301102 -2.158e+00  1.065e+00  -2.025 0.042838 *
factor(GEOID)24025301301 -5.330e-01  8.508e-01  -0.626 0.530997
factor(GEOID)24025301702 -1.663e+01  1.052e+03  -0.016 0.987388
factor(GEOID)24025302100 -1.088e+00  5.825e-01  -1.868 0.061785 .
factor(GEOID)24025302200 -1.455e+00  7.884e-01  -1.846 0.064920 .
factor(GEOID)24025302400 -1.249e+00  8.131e-01  -1.537 0.124386
factor(GEOID)24025303101 -1.369e+00  7.947e-01  -1.723 0.084853 .
factor(GEOID)24025303102 -2.218e-02  9.258e-01  -0.024 0.980890
factor(GEOID)24025303201 -6.283e-01  6.039e-01  -1.040 0.298173
factor(GEOID)24025303204 -1.226e+00  1.106e+00  -1.108 0.267704
factor(GEOID)24025303205 -1.226e+00  1.106e+00  -1.108 0.267704
factor(GEOID)24025303300 -6.615e-01  5.514e-01  -1.200 0.230299
factor(GEOID)24025303400 -1.797e+00  1.076e+00  -1.670 0.094838 .
factor(GEOID)24025303603 -1.326e+00  1.140e+00  -1.163 0.245009
factor(GEOID)24025303700 -1.814e+00  1.065e+00  -1.703 0.088627 .
factor(GEOID)24025304101 -1.315e+00  5.230e-01  -2.514 0.011953 *
factor(GEOID)24025304102 -1.199e+00  4.787e-01  -2.505 0.012244 *
factor(GEOID)24025304201 -1.263e+00  5.701e-01  -2.215 0.026749 *
factor(GEOID)24025304202 -1.250e+00  5.734e-01  -2.181 0.029217 *
factor(GEOID)24025305100 -1.613e+00  4.365e-01  -3.695 0.000220 ***
factor(GEOID)24025305200 -1.417e-01  5.342e-01  -0.265 0.790840
factor(GEOID)24025305300 -1.407e-02  6.083e-01  -0.023 0.981550
factor(GEOID)24025306300 -1.326e+00  1.140e+00  -1.163 0.245009
factor(GEOID)24027601203 -1.618e+01  1.073e+03  -0.015 0.987968
factor(GEOID)24027602202 -4.643e-01  9.388e-01  -0.495 0.620892
factor(GEOID)24027603003 -7.739e-01  5.472e-01  -1.414 0.157293
factor(GEOID)24027603004 -6.283e-01  6.039e-01  -1.040 0.298173
factor(GEOID)24027604001 -1.105e+00  4.868e-01  -2.269 0.023266 *
factor(GEOID)24027604002 -1.453e+00  4.781e-01  -3.040 0.002365 **
factor(GEOID)24027605102 -4.052e-01  5.610e-01  -0.722 0.470169
factor(GEOID)24027605103 -1.891e+00  1.078e+00  -1.755 0.079293 .
factor(GEOID)24027605104 -2.314e+00  1.052e+00  -2.199 0.027904 *
factor(GEOID)24027605602 -1.003e+00  1.129e+00  -0.889 0.374173
factor(GEOID)24027606603 -1.369e+00  1.113e+00  -1.230 0.218675
factor(GEOID)24027606806 -1.630e+01  8.970e+02  -0.018 0.985502
factor(GEOID)24029950100  9.886e-04  2.923e-01   0.003 0.997301
factor(GEOID)24029950200 -2.487e-02  2.595e-01  -0.096 0.923656
factor(GEOID)24029950300  2.433e-01  6.570e-01   0.370 0.711129
factor(GEOID)24029950400 -2.230e-02  2.880e-01  -0.077 0.938277
factor(GEOID)24029950500 -3.020e-01  3.544e-01  -0.852 0.394168
factor(GEOID)24031700101 -1.052e+00  8.237e-01  -1.277 0.201454
factor(GEOID)24031700103 -1.420e+00  5.111e-01  -2.778 0.005476 **
```

```
factor(GEOID)24031700204 -2.166e+00  1.060e+00  -2.044 0.040962 *
factor(GEOID)24031700205 -1.618e+01  7.998e+02  -0.020 0.983858
factor(GEOID)24031700206 -2.215e-01  7.514e-01  -0.295 0.768127
factor(GEOID)24031700208 -1.634e+01  1.058e+03  -0.015 0.987674
factor(GEOID)24031700312 -1.956e+00  7.613e-01  -2.569 0.010200 *
factor(GEOID)24031700400 -1.508e+00  4.702e-01  -3.208 0.001336 **
factor(GEOID)24031700500 -1.732e+00  4.319e-01  -4.009 6.10e-05 ***
factor(GEOID)24031700604 -1.427e+00  6.514e-01  -2.191 0.028427 *
factor(GEOID)24031700608 -1.534e+00  1.097e+00  -1.399 0.161920
factor(GEOID)24031701205 -1.326e+00  1.140e+00  -1.163 0.245009
factor(GEOID)24031701206 -1.630e+01  8.970e+02  -0.018 0.985502
factor(GEOID)24031701307 -1.134e-01  7.889e-01  -0.144 0.885734
factor(GEOID)24031701316 -1.326e+00  1.140e+00  -1.163 0.245009
factor(GEOID)24031701317 -2.500e+00  1.045e+00  -2.393 0.016698 *
factor(GEOID)24031701407 -1.170e+00  1.138e+00  -1.029 0.303635
factor(GEOID)24031701408 -1.607e+00  7.786e-01  -2.064 0.039058 *
factor(GEOID)24031706005 -1.651e+00  1.099e+00  -1.502 0.133140
factor(GEOID)24031706008 -1.705e-01  9.351e-01  -0.182 0.855334
factor(GEOID)24033800411 -1.919e+00  1.060e+00  -1.810 0.070223 .
factor(GEOID)24033800504 -1.503e+00  1.116e+00  -1.346 0.178146
factor(GEOID)24033800507 -8.292e-03  5.657e-01  -0.015 0.988304
factor(GEOID)24033800517 -1.618e+01  1.073e+03  -0.015 0.987968
factor(GEOID)24033800518 -1.632e+01  9.675e+02  -0.017 0.986543
factor(GEOID)24033800605 -1.139e+00  8.251e-01  -1.380 0.167511
factor(GEOID)24033800607 -1.992e+00  1.070e+00  -1.861 0.062780 .
factor(GEOID)24033800701 -9.421e-01  6.111e-01  -1.542 0.123122
factor(GEOID)24033800705 -4.643e-01  9.388e-01  -0.495 0.620892
factor(GEOID)24033800800 -8.510e-01  4.174e-01  -2.039 0.041472 *
factor(GEOID)24033800900 -9.594e-02  3.745e-01  -0.256 0.797820
factor(GEOID)24033801003 -9.611e-02  4.842e-01  -0.198 0.842676
factor(GEOID)24033801004 -2.504e+00  1.042e+00  -2.403 0.016247 *
factor(GEOID)24033801302 -8.207e-01  6.761e-01  -1.214 0.224832
factor(GEOID)24033801305 -1.369e+00  1.113e+00  -1.230 0.218675
factor(GEOID)24033801307 -8.064e-01  8.898e-01  -0.906 0.364806
factor(GEOID)24033801310 -1.992e+00  1.070e+00  -1.861 0.062780 .
factor(GEOID)24033801311 -1.618e+01  9.069e+02  -0.018 0.985764
factor(GEOID)24033801312  1.651e+00  1.134e+00   1.456 0.145508
factor(GEOID)24033802201 -1.170e+00  1.138e+00  -1.029 0.303635
factor(GEOID)24033807301 -1.226e+00  1.106e+00  -1.108 0.267704
factor(GEOID)24033807408 -1.634e+01  1.058e+03  -0.015 0.987674
factor(GEOID)24035810100  1.332e-01  3.805e-01   0.350 0.726218
factor(GEOID)24035810200 -4.218e-01  3.377e-01  -1.249 0.211627
factor(GEOID)24035810300 -4.218e-01  3.377e-01  -1.249 0.211627
```

```
factor(GEOID)24035810400 -1.674e+00  4.340e-01  -3.858 0.000115 ***
factor(GEOID)24035810500 -1.037e+00  3.055e-01  -3.396 0.000684 ***
factor(GEOID)24035810600 -1.399e+00  3.785e-01  -3.696 0.000219 ***
factor(GEOID)24035810700 -1.651e+01  8.882e+02  -0.019 0.985169
factor(GEOID)24035810901 -1.472e+00  7.815e-01  -1.884 0.059567 .
factor(GEOID)24035810902 -1.534e+00  1.097e+00  -1.399 0.161920
factor(GEOID)24035811000  8.258e-01  7.163e-01   1.153 0.248922
factor(GEOID)24037875000 -2.654e+00  1.041e+00  -2.548 0.010820 *
factor(GEOID)24037875100 -1.288e+00  5.855e-01  -2.200 0.027823 *
factor(GEOID)24037875201 -2.442e+00  7.481e-01  -3.264 0.001100 **
factor(GEOID)24037875202 -7.276e-02  3.924e-01  -0.185 0.852889
factor(GEOID)24037875300 -2.396e-01  3.762e-01  -0.637 0.524120
factor(GEOID)24037875400 -1.440e+00  4.706e-01  -3.059 0.002219 **
factor(GEOID)24037875500 -1.209e+00  4.323e-01  -2.796 0.005167 **
factor(GEOID)24037875600 -5.120e-01  4.808e-01  -1.065 0.286909
factor(GEOID)24037875700 -1.788e+00  5.568e-01  -3.211 0.001322 **
factor(GEOID)24037875802 -9.623e-01  8.209e-01  -1.172 0.241114
factor(GEOID)24037875901 -8.695e-01  8.165e-01  -1.065 0.286936
factor(GEOID)24037875902 -3.099e-01  8.797e-01  -0.352 0.724657
factor(GEOID)24037876002  2.847e-01  5.468e-01   0.521 0.602552
factor(GEOID)24037876100 -1.691e+00  5.538e-01  -3.053 0.002265 **
factor(GEOID)24037876200 -8.287e-01  4.171e-01  -1.987 0.046910 *
factor(GEOID)24039930101 -3.512e-01  3.623e-01  -0.969 0.332409
factor(GEOID)24039930102  7.350e-02  3.715e-01   0.198 0.843142
factor(GEOID)24039930200 -2.516e-01  3.152e-01  -0.798 0.424857
factor(GEOID)24039930300 -5.023e-02  2.515e-01  -0.200 0.841694
factor(GEOID)24039930500 -3.133e-02  2.797e-01  -0.112 0.910805
factor(GEOID)24039930600 -1.696e+00  1.072e+00  -1.582 0.113541
factor(GEOID)24041960100 -5.014e-03  2.611e-01  -0.019 0.984679
factor(GEOID)24041960201 -1.025e+00  3.591e-01  -2.856 0.004295 **
factor(GEOID)24041960501  6.103e-02  4.132e-01   0.148 0.882569
factor(GEOID)24041960502 -1.369e+00  6.549e-01  -2.091 0.036529 *
factor(GEOID)24041960600 -1.915e+00  6.264e-01  -3.057 0.002238 **
factor(GEOID)24041960700 -1.323e+00  5.749e-01  -2.302 0.021330 *
factor(GEOID)24041960800 -1.667e+00  6.414e-01  -2.599 0.009360 **
factor(GEOID)24041960900 -1.594e+00  3.950e-01  -4.035 5.46e-05 ***
factor(GEOID)24043010200 -3.620e-01  4.092e-01  -0.885 0.376356
factor(GEOID)24043010300 -7.838e-01  6.873e-01  -1.140 0.254125
factor(GEOID)24043010400 -3.679e-01  5.570e-01  -0.660 0.508988
factor(GEOID)24043010500  7.686e-01  4.097e-01   1.876 0.060676 .
factor(GEOID)24043010600  1.975e-01  2.667e-01   0.741 0.458918
factor(GEOID)24043010700 -2.573e-01  3.083e-01  -0.834 0.404084
factor(GEOID)24043010801  2.964e-01  7.283e-01   0.407 0.683983
```

```
factor(GEOID)24043010802 -3.141e-01  6.512e-01  -0.482 0.629603
factor(GEOID)24043010900 -3.879e-01  3.627e-01  -1.069 0.284946
factor(GEOID)24043011100 -1.705e-01  6.701e-01  -0.254 0.799173
factor(GEOID)24043011202 -4.386e-01  6.367e-01  -0.689 0.490951
factor(GEOID)24043011301 -1.515e-01  5.104e-01  -0.297 0.766623
factor(GEOID)24043011302 -1.404e+00  5.729e-01  -2.450 0.014274 *
factor(GEOID)24043011400  4.205e-01  5.331e-01   0.789 0.430212
factor(GEOID)24043011500 -2.375e+00  5.431e-01  -4.372 1.23e-05 ***
factor(GEOID)24043011600 -7.361e-01  5.178e-01  -1.422 0.155131
factor(GEOID)24045010101 -1.222e+00  8.253e-01  -1.480 0.138754
factor(GEOID)24045010300 -9.169e-01  6.910e-01  -1.327 0.184545
factor(GEOID)24045010400 -3.827e-01  6.531e-01  -0.586 0.557864
factor(GEOID)24045010603 -1.504e-01  4.095e-01  -0.367 0.713459
factor(GEOID)24045010604 -7.046e-01  4.001e-01  -1.761 0.078252 .
factor(GEOID)24045010605 -2.201e+00  7.576e-01  -2.906 0.003663 **
factor(GEOID)24045010606 -1.105e-01  3.079e-01  -0.359 0.719733
factor(GEOID)24045010701  2.473e-01  3.329e-01   0.743 0.457565
factor(GEOID)24045010702 -3.624e-01  4.008e-01  -0.904 0.365910
factor(GEOID)24045010800 -3.985e-02  2.530e-01  -0.157 0.874860
factor(GEOID)24047950100 -4.386e-01  8.872e-01  -0.494 0.621081
factor(GEOID)24047950400 -1.534e+00  1.097e+00  -1.399 0.161920
factor(GEOID)24047950700 -1.170e+00  1.138e+00  -1.029 0.303635
factor(GEOID)24047950800 -2.093e-01  2.679e-01  -0.781 0.434651
factor(GEOID)24047950900  5.871e-02  4.012e-01   0.146 0.883652
factor(GEOID)24047951100 -1.630e+01  8.970e+02  -0.018 0.985502
factor(GEOID)24047951200 -2.180e-01  2.371e-01  -0.919 0.357923
factor(GEOID)24047951300  5.757e-01  7.842e-01   0.734 0.462914
factor(GEOID)24047951400 -2.769e-01  2.307e-01  -1.200 0.230005
factor(GEOID)24047951500  7.050e-02  4.588e-01   0.154 0.877872
factor(GEOID)24047951700 -8.141e-01  5.080e-01  -1.603 0.109044
factor(GEOID)24510260404 -3.189e-01  9.394e-01  -0.339 0.734234
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 10521.9  on 8455  degrees of freedom
Residual deviance:  9407.5  on 8109  degrees of freedom
AIC: 10102

Number of Fisher Scoring iterations: 15
```

Target Odds Ratio: 2

Odds Ratios:

(Intercept)

0.681616524751589

exposure

2.04268762631616

factor(GEOID)24001000200

0.952117194610293

factor(GEOID)24001001300

2.63984615354253

factor(GEOID)24001001401

2.63984615354252

factor(GEOID)24001001502

1.58606415116011

factor(GEOID)24001001503

0.46137967164461

factor(GEOID)24001001600

0.675579880632134

factor(GEOID)24001001700

1.38491108887379

factor(GEOID)24001001900

0.523465279421297

factor(GEOID)24001002000

0.648523835711707

factor(GEOID)24001002100

0.602015878291616

factor(GEOID)24001002200

0.727989860406689

factor(GEOID)24003701200

9.37362956262913e-08

factor(GEOID)24003701300

0.3072304839287

factor(GEOID)24003701400

0.0589414209794206

factor(GEOID)24003702100

0.115569881215388

factor(GEOID)24003702402

0.172127703898875

factor(GEOID)24003702500

0.843259089232406

factor(GEOID)24003702702

0.921200452433772

factor(GEOID)24003707001

0.323780619465672

factor(GEOID)24003708001

0.114632020595044

factor(GEOID)24003730100

0.125803762293105

factor(GEOID)24003730902

0.729334458461488

factor(GEOID)24003731202

0.53720049875909

factor(GEOID)24003731203

8.18693099588413e-08

factor(GEOID)24003731303

0.733550290879791

factor(GEOID)24003731306

7.22973474463093e-08

factor(GEOID)24003740500

7.98020588392534e-08

factor(GEOID)24003740702

0.0989081864298369

factor(GEOID)24003740800

9.37362956263519e-08

factor(GEOID)24003740900

9.37362956263819e-08

factor(GEOID)24003751200

0.771500919771986

factor(GEOID)24003751500

0.366775145439897

factor(GEOID)24003751600

1.23065341003165

factor(GEOID)24003751700

0.226330920853908

factor(GEOID)24005402201

0.153170456845585

factor(GEOID)24005402202

9.37362956263795e-08

factor(GEOID)24005403803

0.293420116351916

factor(GEOID)24005404101

8.34004621222999e-08

factor(GEOID)24005404402

0.366775145439896

factor(GEOID)24005404600

0.171365843220263

factor(GEOID)24005404800

0.136443321133932

factor(GEOID)24005404900

0.335137191776593

factor(GEOID)24005405000

0.301826221890561

factor(GEOID)24005406000

0.15806007360014

factor(GEOID)24005407002

0.204504888676156

factor(GEOID)24005408100

0.586840232703834

factor(GEOID)24005408200

0.3582041910716

factor(GEOID)24005408900

0.310292669657129

factor(GEOID)24005410100

0.404152487604662

factor(GEOID)24005410200

0.21569062220257

factor(GEOID)24005411101

7.22973474462939e-08

factor(GEOID)24005411102

0.244516763626598

factor(GEOID)24005411202

0.254260387401638

factor(GEOID)24005411302

0.254260387401638

factor(GEOID)24005451701

0.420832396410423

factor(GEOID)24005451801

1.67638465437496

factor(GEOID)24005451900

2.63984615354253

factor(GEOID)24005490100

9.3736295626258e-08

factor(GEOID)24009860101

0.1069099587077

factor(GEOID)24009860102

0.191885781563734

factor(GEOID)24009860200

0.286662863694257

factor(GEOID)24009860300

0.871941688418405

factor(GEOID)24009860401

0.259725807233073

factor(GEOID)24009860402

5.99791302906857

factor(GEOID)24009860501

1.39645303758082

factor(GEOID)24009860502

9.37362956265078e-08

factor(GEOID)24009860600

0.0989081864298369

factor(GEOID)24009860701

0.380898715393582

factor(GEOID)24009860702

0.475579105857156

factor(GEOID)24009860703

0.43026096567885

factor(GEOID)24009860801

0.296001461174138

factor(GEOID)24009860802

0.628757392182678

factor(GEOID)24009860900

0.434747635671064

factor(GEOID)24009861001

0.366775145439895

factor(GEOID)24011955000

0.489689062941642

factor(GEOID)24011955100

1.39883629772355

factor(GEOID)24011955201

1.26062302569414

factor(GEOID)24011955202

0.513315795248983

factor(GEOID)24011955301

1.13775782674282

factor(GEOID)24011955400

0.493424213621008

factor(GEOID)24011955500

0.857326491793396

factor(GEOID)24011955600

0.802192517110835

factor(GEOID)24013501001

0.468570702083541

factor(GEOID)24013501002

0.941053395161365

factor(GEOID)24013502000

0.113727386097149

factor(GEOID)24013503000

0.71714072555766

factor(GEOID)24013504100

0.212552760773763

factor(GEOID)24013504201

0.382025225330298

factor(GEOID)24013504202

0.3256249541074

factor(GEOID)24013505102

8.18693099588195e-08

factor(GEOID)24013505206

9.37362956267795e-08

factor(GEOID)24013506101

0.450542448352158

factor(GEOID)24013506102

0.673164505257241

factor(GEOID)24013506200

0.063905087205681

factor(GEOID)24013507500

0.801278872287652

factor(GEOID)24013507601

0.568168754101515

factor(GEOID)24013507602

1.57662474143454

factor(GEOID)24013507702

0.730456739415708

factor(GEOID)24013507802

0.399772003693336

factor(GEOID)24013508101

0.106909958707701

factor(GEOID)24013508200

0.543814317966514

factor(GEOID)24013509001

0.442405434565487

factor(GEOID)24013509002

0.782793047211976

factor(GEOID)24013510000

0.639857513461001

factor(GEOID)24013511000

0.185371265533527

factor(GEOID)24013513002

0.335137191776593

factor(GEOID)24013514201

0.35801125441742

factor(GEOID)24015030100

0.759595273252696

factor(GEOID)24015030200

0.491645362113798

factor(GEOID)24015030501

0.728400450487577

factor(GEOID)24015030503

0.449132669122826

factor(GEOID)24015030505

0.644970389845386

factor(GEOID)24015030601

0.644970389845387

factor(GEOID)24015030602

0.293420116351915

factor(GEOID)24015030700

0.562311462322685

factor(GEOID)24015030903

1.08360623468533

factor(GEOID)24015030904

0.282950683276257

factor(GEOID)24015030905

0.867895455744206

factor(GEOID)24015030906

0.782671440617187

factor(GEOID)24015031201

0.754293026604197

factor(GEOID)24015031202

0.729334458461489

factor(GEOID)24015031301

0.502379952280847

factor(GEOID)24015031302

0.688102130459171

factor(GEOID)24015031400

1.21071429681918

factor(GEOID)24017850101

0.908365192869896

factor(GEOID)24017850102

0.849112258412634

factor(GEOID)24017850201

0.851321491922113

factor(GEOID)24017850300

0.682017950069052

factor(GEOID)24017850400

0.887995296058819

factor(GEOID)24017850500

0.249906793316277

factor(GEOID)24017850600

0.163334899646813

factor(GEOID)24017850712

0.441516604617526

factor(GEOID)24017850713

7.22973474463314e-08

factor(GEOID)24017850802

0.933221316054675

factor(GEOID)24017850904

0.168857006563655

factor(GEOID)24017851001

0.254260387401638

factor(GEOID)24017851002

8.92169454520618e-08

factor(GEOID)24017851100

0.837348600686485

factor(GEOID)24017851200

1.0471430114204

factor(GEOID)24017851301

0.250751266597906

factor(GEOID)24017851302

0.493009337108679

factor(GEOID)24017851400

0.264616351427399

factor(GEOID)24017851500

7.98020588390419e-08

factor(GEOID)24019970100

0.947255891982318

factor(GEOID)24019970200

1.03075769908115

factor(GEOID)24019970300

0.906346714690158

factor(GEOID)24019970600

0.729334458461488

factor(GEOID)24019970702

1.25538805429552

factor(GEOID)24019970804

1.43741685788276

factor(GEOID)24019970900

0.871066425680801

factor(GEOID)24021740200

0.46857070208354

factor(GEOID)24021751003

0.265669171484409

factor(GEOID)24021751004

0.293420116351916

factor(GEOID)24021751301

0.165828376677913

factor(GEOID)24021751302

0.626236203043281

factor(GEOID)24021751600

1.45609252599698

factor(GEOID)24021751701

0.310292669657129

factor(GEOID)24021751702

0.101015407459997

factor(GEOID)24021751801

0.554222200381026

factor(GEOID)24021751802

0.119922423964362

factor(GEOID)24021751902

0.31029266965713

factor(GEOID)24021751904

0.803719217245562

factor(GEOID)24021752001

0.978067054506386

factor(GEOID)24021752102

0.170459931247324

factor(GEOID)24021752201

0.154805394186044

factor(GEOID)24021752204

9.3736295626615e-08

factor(GEOID)24021752302

0.411140033883256

factor(GEOID)24021752303

0.991741846728106

factor(GEOID)24021752501

0.518570099558633

factor(GEOID)24021752502

9.3736295626615e-08

factor(GEOID)24021752601

7.98020588390688e-08

factor(GEOID)24021752602

0.323780619465672

factor(GEOID)24021752801

0.559846396295156

factor(GEOID)24021752802

0.254260387401638

factor(GEOID)24021752900

1.03598297600151

factor(GEOID)24021766800

2.93420116351915

factor(GEOID)24021767500

0.816478462636148

factor(GEOID)24021767600

0.304120345138845

factor(GEOID)24021770700

0.423733989235644

factor(GEOID)24021775302

0.434747635671063

factor(GEOID)24023000200

0.773008151355644

factor(GEOID)24023000300

0.661348317285466

factor(GEOID)24023000400

0.682017950069051

factor(GEOID)24023000500

0.251577636300338

factor(GEOID)24023000600

0.796683248350836

factor(GEOID)24023000700

1.02661580438856

factor(GEOID)24025301102

0.115569881215388

factor(GEOID)24025301301

0.586840232703834

factor(GEOID)24025301702

5.9998784451814e-08

factor(GEOID)24025302100

0.336886308995621

factor(GEOID)24025302200

0.233355033853536

factor(GEOID)24025302400

0.286662863694257

factor(GEOID)24025303101

0.254260387401637

factor(GEOID)24025303102

0.97806705450639

factor(GEOID)24025303201

0.533491120639847

factor(GEOID)24025303204

0.293420116351917

factor(GEOID)24025303205

0.293420116351915

factor(GEOID)24025303300

0.516096538251043

factor(GEOID)24025303400

0.165828376677912

factor(GEOID)24025303603

0.265669171484408

factor(GEOID)24025303700

0.163011175751065

factor(GEOID)24025304101

0.268578897681667

factor(GEOID)24025304102

0.301424396798773

factor(GEOID)24025304201

0.282840661465373

factor(GEOID)24025304202

0.286390876253849

factor(GEOID)24025305100

0.199297121489622

factor(GEOID)24025305200

0.867895455744204

factor(GEOID)24025305300

0.986031775735797

factor(GEOID)24025306300

0.265669171484408

factor(GEOID)24027601203

9.37362956262677e-08

factor(GEOID)24027602202

0.628556242330117

factor(GEOID)24027603003

0.461211758681117

factor(GEOID)24027603004

0.533491120639848

factor(GEOID)24027604001

0.331328175894313

factor(GEOID)24027604002

0.23378228466624

factor(GEOID)24027605102

0.66686390079981

factor(GEOID)24027605103

0.150884499444064

factor(GEOID)24027605104

0.0989081864298362

factor(GEOID)24027605602

0.366775145439896

factor(GEOID)24027606603

0.254260387401638

factor(GEOID)24027606806

8.34004621225506e-08

factor(GEOID)24029950100

1.00098912697541

factor(GEOID)24029950200

0.975437352231715

factor(GEOID)24029950300

1.27546171823109

factor(GEOID)24029950400

0.977946764067791

factor(GEOID)24029950500

0.739322892652372

factor(GEOID)24031700101

0.349177147602124

factor(GEOID)24031700103

0.241774547897694

factor(GEOID)24031700204

0.114632020595043

factor(GEOID)24031700205

9.37362956264834e-08

factor(GEOID)24031700206

0.801278872287654

factor(GEOID)24031700208

7.98020588379172e-08

factor(GEOID)24031700312

0.141460274154228

factor(GEOID)24031700400

0.221267177849158

factor(GEOID)24031700500

0.177005404147658

factor(GEOID)24031700604

0.239921099762789

factor(GEOID)24031700608

0.215690622202568

factor(GEOID)24031701205

0.265669171484409

factor(GEOID)24031701206

8.34004621224347e-08

factor(GEOID)24031701307

0.892815640952924

factor(GEOID)24031701316

0.265669171484408

factor(GEOID)24031701317

0.082091775077708

factor(GEOID)24031701407

0.310292669657129

factor(GEOID)24031701408

0.200558335740784

factor(GEOID)24031706005

0.191885781563734

factor(GEOID)24031706008

0.843259089232404

factor(GEOID)24033800411

0.146710058175958

factor(GEOID)24033800504

0.222556195524394

factor(GEOID)24033800507

0.991741846728105

factor(GEOID)24033800517

9.3736295626579e-08

factor(GEOID)24033800518

8.18693099589184e-08

factor(GEOID)24033800605

0.320205110737331

factor(GEOID)24033800607

0.136443321133932

factor(GEOID)24033800701

0.389790723375907

factor(GEOID)24033800705

0.628556242330116

factor(GEOID)24033800800

0.426985357650946

factor(GEOID)24033800900

0.908515106018458

factor(GEOID)24033801003

0.908365192869896

factor(GEOID)24033801004

0.0817960538637789

factor(GEOID)24033801302

0.440130174527874

factor(GEOID)24033801305

0.254260387401638

factor(GEOID)24033801307

0.446474525226718

factor(GEOID)24033801310

0.136443321133932

factor(GEOID)24033801311

9.37362956264748e-08

factor(GEOID)24033801312

5.21024759304288

factor(GEOID)24033802201

0.310292669657128

factor(GEOID)24033807301

0.293420116351916

factor(GEOID)24033807408

7.98020588387065e-08

factor(GEOID)24035810100

1.14252007390247

factor(GEOID)24035810200

0.655847723375514

factor(GEOID)24035810300

0.655847723375515

factor(GEOID)24035810400

0.187443751502492

factor(GEOID)24035810500

0.354361979346119

factor(GEOID)24035810600

0.246854195403539

factor(GEOID)24035810700

6.75518565702698e-08

factor(GEOID)24035810901

0.229408917112289

factor(GEOID)24035810902

0.215690622202569

factor(GEOID)24035811000

2.28380681640746

factor(GEOID)24037875000

0.0703681254546557

factor(GEOID)24037875100

0.275856440713961

factor(GEOID)24037875201

0.0870165236527611

factor(GEOID)24037875202

0.929820237800668

factor(GEOID)24037875300

0.786910188910703

factor(GEOID)24037875400

0.237038081992992

factor(GEOID)24037875500

0.298560737239839

factor(GEOID)24037875600

0.59927706630639

factor(GEOID)24037875700

0.167314490553396

factor(GEOID)24037875802

0.3820252253303

factor(GEOID)24037875901

0.419171594788452

factor(GEOID)24037875902

0.733550290879792

factor(GEOID)24037876002

1.329398151498

factor(GEOID)24037876100

0.184368751399966

factor(GEOID)24037876200

0.436602417525719

factor(GEOID)24039930101

0.703844710986281

factor(GEOID)24039930102

1.07627348243732

factor(GEOID)24039930200

0.777580613746388

factor(GEOID)24039930300

0.95101338283132

factor(GEOID)24039930500

0.969151084952093

factor(GEOID)24039930600

0.183387572719948

factor(GEOID)24041960100

0.994998392225873

factor(GEOID)24041960201

0.358658625740334

factor(GEOID)24041960501

1.06293097971624

factor(GEOID)24041960502

0.254260387401638

factor(GEOID)24041960600

0.147372429457199

factor(GEOID)24041960700

0.266227116749581

factor(GEOID)24041960800

0.188878371427332

factor(GEOID)24041960900

0.203121294698429

factor(GEOID)24043010200

0.69631308599127

factor(GEOID)24043010300

0.45669007881154

factor(GEOID)24043010400

0.692207347691557

factor(GEOID)24043010500

2.15673309770052

factor(GEOID)24043010600

1.21838075533814

factor(GEOID)24043010700

0.773165358543001

factor(GEOID)24043010801

1.34505149315709

factor(GEOID)24043010802

0.730456739415711

factor(GEOID)24043010900

0.678500407512681

factor(GEOID)24043011100

0.843259089232405

factor(GEOID)24043011202

0.644970389845387

factor(GEOID)24043011301

0.859435680222981

factor(GEOID)24043011302

0.245698607619209

factor(GEOID)24043011400

1.52279227249689

factor(GEOID)24043011500

0.0930350270582214

factor(GEOID)24043011600

0.478954936979507

factor(GEOID)24045010101

0.294676062164143

factor(GEOID)24045010300

0.399772003693335

factor(GEOID)24045010400

0.682017950069052

factor(GEOID)24045010603

0.860388029216669

factor(GEOID)24045010604

0.494328721918774

factor(GEOID)24045010605

0.110651659982268

factor(GEOID)24045010606

0.895416175126686

factor(GEOID)24045010701

1.28056448711764

factor(GEOID)24045010702

0.695984176282028

factor(GEOID)24045010800

0.960931356046071

factor(GEOID)24047950100

0.644970389845387

factor(GEOID)24047950400

0.215690622202569

factor(GEOID)24047950700

0.31029266965713

factor(GEOID)24047950800

0.811146082474215

factor(GEOID)24047950900

1.06046894111815

factor(GEOID)24047951100

8.34004621225838e-08

factor(GEOID)24047951200

0.804159010819759

factor(GEOID)24047951300

1.77832161712388

factor(GEOID)24047951400

0.758125591532237

factor(GEOID)24047951500

1.07304667216065

factor(GEOID)24047951700

0.443023898391492

factor(GEOID)24510260404

0.726935160418227

Further Reading

Geostatistics

Geostatistics assess variables across space. For example, do neighboring census tracts have similar values of ndi? This can be used to assess model assumptions like the independence of observations (i.e. clustering).

Moran's I

Moran's I is an example of an overall summary statistic that assess the degree of autocorrelation present in in the data.

```
# install pacakge used in this section
install.packages('spdep')
library(spdep)
```

```
Installing package into '/content/library'
(as 'lib' is unspecified)

Loading required package: spData

To access larger datasets in this package, install the spDataLarge
package with: `install.packages('spDataLarge',
repos='https://nowosad.github.io/drat/', type='source')`
```

```
# create "complete" dataset (fill in ndi values that are missing)
md_complete = md_ndi %>% mutate(NDI = ifelse(is.na(NDI), mean(NDI, na.rm=TRUE), NDI))

# calculate neighbors
nb <- poly2nb(md_complete, queen=TRUE)
# calculate neighbor weights
lw <- nb2listw(nb, style="W", zero.policy=TRUE)
# calculate Moran's I and test statistic
```

```r
moran(md_complete$NDI, lw, length(nb), Szero(lw))
moran.test(md_complete$NDI, lw)
```

I < /*dt* >< *dd* > 0.680114167847662 < /*dd* >< *dt* >K

```
<dd>2.88280081363596</dd>
```

```
	Moran I test under randomisation

data:  md_complete$NDI
weights: lw

Moran I statistic standard deviate = 42.722, p-value < 2.2e-16
alternative hypothesis: greater
sample estimates:
Moran I statistic       Expectation          Variance
     0.6801141678     -0.0007168459      0.0002539641
```

G and Gstar

These statistics are local - you can plot them on a map. They show clustering "hot" and "cold" spots.

```r
md_complete$G = localG(md_complete$NDI, listw=lw)
ggplot(md_complete) +
  geom_sf(aes(fill=as.numeric(G))) +
  scale_fill_gradient2(low = "blue", mid = "white", high = "red", midpoint = 0) +
  labs(fill='G Statistic')
```

Raster Data

So far, we've looked at spatial data stored as point and polygons. Rasters are another format of geospatial data consisting of a complete grid of observations. They are commonly used for data derived from satellite imagery.

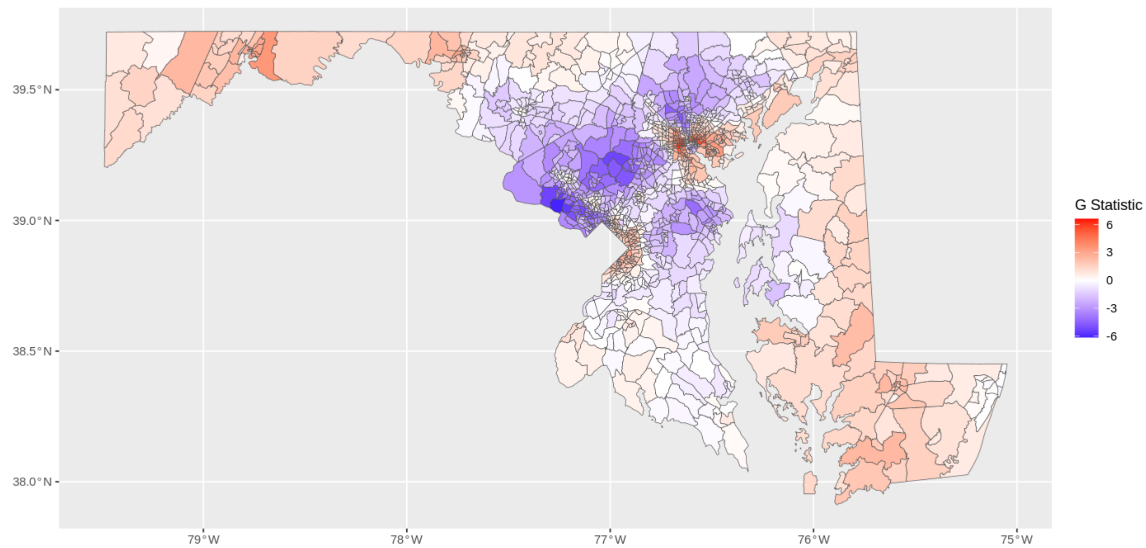

Figure 9: png

Running on NIH HPC (Biowulf)

Since this is a Jupyter notebook, you can run this code on the NIH HPC using the instructions here.

The basic procedure is:

1. Sign up for a Biowulf account.
2. From the login node, run an interactive job (e.g. sinteractive –tunnel).
3. Copy the resulting tunnel and connect to the assigned node.
4. Start Jupyter Lab and navigate to the resulting notebook URL.

Installing from CRAN

If the setup code at the top of the script isn't working, try installing the packages using the code block below. This takes ~20 minutes on Colab. Make sure you do this on a clean R runtime (in the top-right menu, delete and disconnect from the current runtime and then click restart runtime).

```
# create a directory to store our R libraries
system("mkdir library")
.libPaths("library")

# download packages
install.packages('classInt')
install.packages('units')
install.packages("sf")
install.packages("ggplot2")
install.packages("dplyr")
install.packages('tigris')

# test packages
library(sf)
library(ggplot2)
library(dplyr)
library(tigris)
```

Repackage

Use the code in this section to repackage the R libraries to rebuild the workshop in the future. You'll need to host this file somewhere where it can be easily downloaded (like Google Drive).

```
# compress files
system('tar -czvf rlib.tar.gz library')
# then download from the file browser and upload somewhere that allows direct download lin
```

Help!

If you get stuck, try these steps:

- Try reloading the page, sometimes it will time out (google will return your loaned virtual machine to the pool if it isn't being used).
- Make sure you run the code in order, for example you need to run the first code block that installs and loads the R packages to get anything else to run.
- If you edited the code and can't get it to run anymore, just start over in a fresh copy of the tutorial.