Making a reproducible project

Chapter 6 case study

John Little Sophia Lafferty-Hess

2022-11-10

Make a Quarto project

This Quarto project is also a git repo with {renv} dependency management. i.e. ***reproducibility!***

You can and, generally, should set up your Posit/Quart project[1] as a *git* repository. Do this when you create your Quarto document. If you did not, you can retroactively configure your RStudio project as a git-enabled and renv-enabled Quarto Project.[2][3]

We will use Posit/RStudio's native ability to orchestrate git and GitHub. You should create a free GitHub account and configure your RStudio with your GitHub Personal Access Token.[4]

Load library packages

As a general rule for our project we will load eight Tidyverse packages by invoking the following command: `library(tidyverse)`.[5]

This project will import census data in the second code-chunk. The Census ACS data were initially downloaded by the workshop authors via the {tidycensus} package. For this project we will focus on managing existing census data as part of our reproducible computational workflow.[6]

```
library(tidyverse)
library(janitor)
library(fs)
library(here)
library(renv)
library(sf)
#library(tidycensus)
```

[1] Tooltip background on RStudio projects

[2] In this case, the the {usethis} library package is helpful.

[3] To learn more about setting up git with R, please see the quick tips for doing using git and RStudio with the {usethis} package by visiting the Rfun git workshop page.

[4] More information can be found at the {usethis} package web page; See *Managing git(Hub) credentials*. A quick guide to managing git and GitHub credentials can be found at the Rfun git quick reference page and more information can be found at the Rfun git/GitHub workshop page.

[5] You can reduce the overall size of your project by being more selective regarding which tidyverse packages you use. e.g. Are you using only {ggplot2} and {dplyr}? You can load those only those packages, instead of loading all eight of the eight defualt tidyverse packages.

[6] If you want to learn more about {tidycensus}, you can see the tidycensus package website and invoke the `library(tidycensus)`.

Gather data

```
nc_pop <- sf::read_sf(here("data/raw/ACS_nc-county_shapefile.shp"))
nc_pop |>
  as_tibble()
```

GEOID	NAME	variable	estimate	moe	geometry
37039	Cherokee County, North Carolina	B01003_001	28413	NA	MULTIPOLYGON (((-84.31749 3…
37089	Henderson County, North Carolina	B01003_001	116298	NA	MULTIPOLYGON (((-82.74289 3…
37171	Surry County, North Carolina	B01003_001	71904	NA	MULTIPOLYGON (((-80.97364 3…
37131	Northampton County, North Carolina	B01003_001	19672	NA	MULTIPOLYGON (((-77.90008 3…
37177	Tyrrell County, North Carolina	B01003_001	3978	NA	MULTIPOLYGON (((-76.4056 35…
37043	Clay County, North Carolina	B01003_001	11150	NA	MULTIPOLYGON (((-84.00582 3…
37075	Graham County, North Carolina	B01003_001	8501	NA	MULTIPOLYGON (((-84.03815 3…
37031	Carteret County, North Carolina	B01003_001	69301	NA	MULTIPOLYGON (((-76.33014 3…
37007	Anson County, North Carolina	B01003_001	24430	NA	MULTIPOLYGON (((-80.3153 34…
37183	Wake County, North Carolina	B01003_001	1091662	48	MULTIPOLYGON (((-78.9877 35…
37083	Halifax County, North Carolina	B01003_001	50678	NA	MULTIPOLYGON (((-78.00655 3…
37125	Moore County, North Carolina	B01003_001	99263	NA	MULTIPOLYGON (((-79.76796 3…
37065	Edgecombe County, North Carolina	B01003_001	52069	NA	MULTIPOLYGON (((-77.82844 3…
37107	Lenoir County, North Carolina	B01003_001	56350	NA	MULTIPOLYGON (((-77.83109 3…
37193	Wilkes County, North Carolina	B01003_001	68341	48	MULTIPOLYGON (((-81.542 36….
37069	Franklin County, North Carolina	B01003_001	68027	NA	MULTIPOLYGON (((-78.54625 3…
37117	Martin County, North Carolina	B01003_001	22644	NA	MULTIPOLYGON (((-77.40261 3…
37137	Pamlico County, North Carolina	B01003_001	12673	NA	MULTIPOLYGON (((-76.98756 3…

GEOID	NAME	variable	estimate	moe	geometry
37149	Polk County, North Carolina	B01003_001	20682	NA	MULTIPOLYGON (((-82.35921 3…
37175	Transylvania County, North Carolina	B01003_001	34039	NA	MULTIPOLYGON (((-83.05693 3…
37143	Perquimans County, North Carolina	B01003_001	13513	NA	MULTIPOLYGON (((-76.58816 3…
37047	Columbus County, North Carolina	B01003_001	55659	NA	MULTIPOLYGON (((-79.07124 3…
37161	Rutherford County, North Carolina	B01003_001	66741	NA	MULTIPOLYGON (((-82.28053 3…
37155	Robeson County, North Carolina	B01003_001	131656	NA	MULTIPOLYGON (((-79.46156 3…
37133	Onslow County, North Carolina	B01003_001	198377	NA	MULTIPOLYGON (((-77.17131 3…
37113	Macon County, North Carolina	B01003_001	35172	NA	MULTIPOLYGON (((-83.73709 3…
37091	Hertford County, North Carolina	B01003_001	23752	NA	MULTIPOLYGON (((-77.20861 3…
37127	Nash County, North Carolina	B01003_001	94287	NA	MULTIPOLYGON (((-78.2556 35…
37101	Johnston County, North Carolina	B01003_001	203308	NA	MULTIPOLYGON (((-78.7089 35…
37111	McDowell County, North Carolina	B01003_001	45402	NA	MULTIPOLYGON (((-82.29491 3…
37021	Buncombe County, North Carolina	B01003_001	259576	NA	MULTIPOLYGON (((-82.88811 3…
37153	Richmond County, North Carolina	B01003_001	44759	NA	MULTIPOLYGON (((-80.07567 3…
37027	Caldwell County, North Carolina	B01003_001	82056	NA	MULTIPOLYGON (((-81.81052 3…
37067	Forsyth County, North Carolina	B01003_001	378499	NA	MULTIPOLYGON (((-80.51556 3…
37093	Hoke County, North Carolina	B01003_001	54590	NA	MULTIPOLYGON (((-79.45792 3…
37085	Harnett County, North Carolina	B01003_001	134328	48	MULTIPOLYGON (((-79.22186 3…
37087	Haywood County, North Carolina	B01003_001	61862	NA	MULTIPOLYGON (((-83.25611 3…
37123	Montgomery County, North Carolina	B01003_001	27223	NA	MULTIPOLYGON (((-80.18256 3…
37103	Jones County, North Carolina	B01003_001	9453	NA	MULTIPOLYGON (((-77.73078 3…
37167	Stanly County, North Carolina	B01003_001	62050	NA	MULTIPOLYGON (((-80.50617 3…

GEOID	NAME	variable	estimate	moe	geometry
37079	Greene County, North Carolina	B01003_001	20987	NA	MULTIPOLYGON (((-77.82647 3…
37017	Bladen County, North Carolina	B01003_001	33209	NA	MULTIPOLYGON (((-78.902 34….
37013	Beaufort County, North Carolina	B01003_001	47160	NA	MULTIPOLYGON (((-77.19577 3…
37053	Currituck County, North Carolina	B01003_001	27210	NA	MULTIPOLYGON (((-76.3133 36…
37041	Chowan County, North Carolina	B01003_001	13995	NA	MULTIPOLYGON (((-76.72232 3…
37019	Brunswick County, North Carolina	B01003_001	137303	NA	MULTIPOLYGON (((-78.65038 3…
37009	Ashe County, North Carolina	B01003_001	27009	48	MULTIPOLYGON (((-81.74065 3…
37141	Pender County, North Carolina	B01003_001	61891	NA	MULTIPOLYGON (((-78.27223 3…
37051	Cumberland County, North Carolina	B01003_001	334562	NA	MULTIPOLYGON (((-79.11285 3…
37163	Sampson County, North Carolina	B01003_001	63284	NA	MULTIPOLYGON (((-78.67179 3…
37023	Burke County, North Carolina	B01003_001	90148	NA	MULTIPOLYGON (((-81.98334 3…
37035	Catawba County, North Carolina	B01003_001	158507	NA	MULTIPOLYGON (((-81.5354 35…
37037	Chatham County, North Carolina	B01003_001	72853	NA	MULTIPOLYGON (((-79.55503 3…
37099	Jackson County, North Carolina	B01003_001	43435	NA	MULTIPOLYGON (((-83.36373 3…
37055	Dare County, North Carolina	B01003_001	36698	NA	MULTIPOLYGON (((-75.72681 3…
37179	Union County, North Carolina	B01003_001	235767	NA	MULTIPOLYGON (((-80.83802 3…
37095	Hyde County, North Carolina	B01003_001	5089	NA	MULTIPOLYGON (((-76.01528 3…
37029	Camden County, North Carolina	B01003_001	10654	NA	MULTIPOLYGON (((-76.54154 3…
37199	Yancey County, North Carolina	B01003_001	17870	NA	MULTIPOLYGON (((-82.50538 3…
37165	Scotland County, North Carolina	B01003_001	34921	NA	MULTIPOLYGON (((-79.69251 3…
37139	Pasquotank County, North Carolina	B01003_001	39775	NA	MULTIPOLYGON (((-76.49134 3…
37061	Duplin County, North Carolina	B01003_001	58965	NA	MULTIPOLYGON (((-78.19854 3…

GEOID	NAME	variable	estimate	moe	geometry
37173	Swain County, North Carolina	B01003_001	14241	NA	MULTIPOLYGON (((-83.94939 3…
37129	New Hanover County, North Carolina	B01003_001	231448	NA	MULTIPOLYGON (((-78.02992 3…
37115	Madison County, North Carolina	B01003_001	21608	NA	MULTIPOLYGON (((-82.96221 3…
37159	Rowan County, North Carolina	B01003_001	140978	NA	MULTIPOLYGON (((-80.77114 3…
37049	Craven County, North Carolina	B01003_001	102290	NA	MULTIPOLYGON (((-77.47329 3…
37119	Mecklenburg County, North Carolina	B01003_001	1095170	NA	MULTIPOLYGON (((-81.05803 3…
37071	Gaston County, North Carolina	B01003_001	222119	NA	MULTIPOLYGON (((-81.4556 35…
37005	Alleghany County, North Carolina	B01003_001	11085	NA	MULTIPOLYGON (((-81.35326 3…
37169	Stokes County, North Carolina	B01003_001	45688	NA	MULTIPOLYGON (((-80.4502 36…
37033	Caswell County, North Carolina	B01003_001	22619	NA	MULTIPOLYGON (((-79.53085 3…
37003	Alexander County, North Carolina	B01003_001	37271	NA	MULTIPOLYGON (((-81.34359 3…
37121	Mitchell County, North Carolina	B01003_001	14959	NA	MULTIPOLYGON (((-82.41666 3…
37077	Granville County, North Carolina	B01003_001	59823	NA	MULTIPOLYGON (((-78.80729 3…
37185	Warren County, North Carolina	B01003_001	19746	NA	MULTIPOLYGON (((-78.32391 3…
37045	Cleveland County, North Carolina	B01003_001	97765	NA	MULTIPOLYGON (((-81.76811 3…
37059	Davie County, North Carolina	B01003_001	42543	NA	MULTIPOLYGON (((-80.70782 3…
37147	Pitt County, North Carolina	B01003_001	179961	NA	MULTIPOLYGON (((-77.70069 3…
37195	Wilson County, North Carolina	B01003_001	81579	NA	MULTIPOLYGON (((-78.19212 3…
37057	Davidson County, North Carolina	B01003_001	166837	NA	MULTIPOLYGON (((-80.48742 3…
37135	Orange County, North Carolina	B01003_001	146354	NA	MULTIPOLYGON (((-79.26843 3…
37105	Lee County, North Carolina	B01003_001	61083	NA	MULTIPOLYGON (((-79.3599 35…
37073	Gates County, North Carolina	B01003_001	11519	NA	MULTIPOLYGON (((-76.95045 3…

GEOID	NAME	variable	estimate	moe	geometry
37157	Rockingham County, North Carolina	B01003_001	91051	NA	MULTIPOLYGON (((-80.03512 3…
37145	Person County, North Carolina	B01003_001	39561	NA	MULTIPOLYGON (((-79.1533 36…
37151	Randolph County, North Carolina	B01003_001	143460	NA	MULTIPOLYGON (((-80.06655 3…
37181	Vance County, North Carolina	B01003_001	44614	NA	MULTIPOLYGON (((-78.51122 3…
37187	Washington County, North Carolina	B01003_001	11788	NA	MULTIPOLYGON (((-76.84726 3…
37011	Avery County, North Carolina	B01003_001	17510	NA	MULTIPOLYGON (((-82.08094 3…
37081	Guilford County, North Carolina	B01003_001	532956	NA	MULTIPOLYGON (((-80.0466 35…
37109	Lincoln County, North Carolina	B01003_001	84580	NA	MULTIPOLYGON (((-81.53674 3…
37025	Cabarrus County, North Carolina	B01003_001	211605	NA	MULTIPOLYGON (((-80.78709 3…
37191	Wayne County, North Carolina	B01003_001	123785	NA	MULTIPOLYGON (((-78.30437 3…
37189	Watauga County, North Carolina	B01003_001	55669	NA	MULTIPOLYGON (((-81.91811 3…
37097	Iredell County, North Carolina	B01003_001	178853	NA	MULTIPOLYGON (((-81.10951 3…
37001	Alamance County, North Carolina	B01003_001	166144	NA	MULTIPOLYGON (((-79.54205 3…
37015	Bertie County, North Carolina	B01003_001	19081	NA	MULTIPOLYGON (((-77.32787 3…
37063	Durham County, North Carolina	B01003_001	317665	NA	MULTIPOLYGON (((-79.01207 3…
37197	Yadkin County, North Carolina	B01003_001	37589	NA	MULTIPOLYGON (((-80.88125 3…

Clean / Normalize / Wrangle

```
nc_pop |>
  sf::st_drop_geometry() |>
  janitor::clean_names() |>
  separate(name, into = c("county", "state"), sep = ",") |>
  mutate(county = str_remove(county, " County")) |>
  rename(population = estimate)
```

geoid	county	state	variable	population	moe
37039	Cherokee	North Carolina	B01003_001	28413	NA
37089	Henderson	North Carolina	B01003_001	116298	NA
37171	Surry	North Carolina	B01003_001	71904	NA
37131	Northampton	North Carolina	B01003_001	19672	NA
37177	Tyrrell	North Carolina	B01003_001	3978	NA
37043	Clay	North Carolina	B01003_001	11150	NA
37075	Graham	North Carolina	B01003_001	8501	NA
37031	Carteret	North Carolina	B01003_001	69301	NA
37007	Anson	North Carolina	B01003_001	24430	NA
37183	Wake	North Carolina	B01003_001	1091662	48
37083	Halifax	North Carolina	B01003_001	50678	NA
37125	Moore	North Carolina	B01003_001	99263	NA
37065	Edgecombe	North Carolina	B01003_001	52069	NA
37107	Lenoir	North Carolina	B01003_001	56350	NA
37193	Wilkes	North Carolina	B01003_001	68341	48
37069	Franklin	North Carolina	B01003_001	68027	NA
37117	Martin	North Carolina	B01003_001	22644	NA
37137	Pamlico	North Carolina	B01003_001	12673	NA
37149	Polk	North Carolina	B01003_001	20682	NA
37175	Transylvania	North Carolina	B01003_001	34039	NA
37143	Perquimans	North Carolina	B01003_001	13513	NA
37047	Columbus	North Carolina	B01003_001	55659	NA
37161	Rutherford	North Carolina	B01003_001	66741	NA
37155	Robeson	North Carolina	B01003_001	131656	NA
37133	Onslow	North Carolina	B01003_001	198377	NA
37113	Macon	North Carolina	B01003_001	35172	NA
37091	Hertford	North Carolina	B01003_001	23752	NA
37127	Nash	North Carolina	B01003_001	94287	NA
37101	Johnston	North Carolina	B01003_001	203308	NA
37111	McDowell	North Carolina	B01003_001	45402	NA
37021	Buncombe	North Carolina	B01003_001	259576	NA
37153	Richmond	North Carolina	B01003_001	44759	NA
37027	Caldwell	North Carolina	B01003_001	82056	NA
37067	Forsyth	North Carolina	B01003_001	378499	NA
37093	Hoke	North Carolina	B01003_001	54590	NA
37085	Harnett	North Carolina	B01003_001	134328	48
37087	Haywood	North Carolina	B01003_001	61862	NA
37123	Montgomery	North Carolina	B01003_001	27223	NA
37103	Jones	North Carolina	B01003_001	9453	NA
37167	Stanly	North Carolina	B01003_001	62050	NA
37079	Greene	North Carolina	B01003_001	20987	NA
37017	Bladen	North Carolina	B01003_001	33209	NA
37013	Beaufort	North Carolina	B01003_001	47160	NA
37053	Currituck	North Carolina	B01003_001	27210	NA

geoid	county	state	variable	population	moe
37041	Chowan	North Carolina	B01003_001	13995	NA
37019	Brunswick	North Carolina	B01003_001	137303	NA
37009	Ashe	North Carolina	B01003_001	27009	48
37141	Pender	North Carolina	B01003_001	61891	NA
37051	Cumberland	North Carolina	B01003_001	334562	NA
37163	Sampson	North Carolina	B01003_001	63284	NA
37023	Burke	North Carolina	B01003_001	90148	NA
37035	Catawba	North Carolina	B01003_001	158507	NA
37037	Chatham	North Carolina	B01003_001	72853	NA
37099	Jackson	North Carolina	B01003_001	43435	NA
37055	Dare	North Carolina	B01003_001	36698	NA
37179	Union	North Carolina	B01003_001	235767	NA
37095	Hyde	North Carolina	B01003_001	5089	NA
37029	Camden	North Carolina	B01003_001	10654	NA
37199	Yancey	North Carolina	B01003_001	17870	NA
37165	Scotland	North Carolina	B01003_001	34921	NA
37139	Pasquotank	North Carolina	B01003_001	39775	NA
37061	Duplin	North Carolina	B01003_001	58965	NA
37173	Swain	North Carolina	B01003_001	14241	NA
37129	New Hanover	North Carolina	B01003_001	231448	NA
37115	Madison	North Carolina	B01003_001	21608	NA
37159	Rowan	North Carolina	B01003_001	140978	NA
37049	Craven	North Carolina	B01003_001	102290	NA
37119	Mecklenburg	North Carolina	B01003_001	1095170	NA
37071	Gaston	North Carolina	B01003_001	222119	NA
37005	Alleghany	North Carolina	B01003_001	11085	NA
37169	Stokes	North Carolina	B01003_001	45688	NA
37033	Caswell	North Carolina	B01003_001	22619	NA
37003	Alexander	North Carolina	B01003_001	37271	NA
37121	Mitchell	North Carolina	B01003_001	14959	NA
37077	Granville	North Carolina	B01003_001	59823	NA
37185	Warren	North Carolina	B01003_001	19746	NA
37045	Cleveland	North Carolina	B01003_001	97765	NA
37059	Davie	North Carolina	B01003_001	42543	NA
37147	Pitt	North Carolina	B01003_001	179961	NA
37195	Wilson	North Carolina	B01003_001	81579	NA
37057	Davidson	North Carolina	B01003_001	166837	NA
37135	Orange	North Carolina	B01003_001	146354	NA
37105	Lee	North Carolina	B01003_001	61083	NA
37073	Gates	North Carolina	B01003_001	11519	NA
37157	Rockingham	North Carolina	B01003_001	91051	NA
37145	Person	North Carolina	B01003_001	39561	NA
37151	Randolph	North Carolina	B01003_001	143460	NA
37181	Vance	North Carolina	B01003_001	44614	NA

geoid	county	state	variable	population	moe
37187	Washington	North Carolina	B01003_001	11788	NA
37011	Avery	North Carolina	B01003_001	17510	NA
37081	Guilford	North Carolina	B01003_001	532956	NA
37109	Lincoln	North Carolina	B01003_001	84580	NA
37025	Cabarrus	North Carolina	B01003_001	211605	NA
37191	Wayne	North Carolina	B01003_001	123785	NA
37189	Watauga	North Carolina	B01003_001	55669	NA
37097	Iredell	North Carolina	B01003_001	178853	NA
37001	Alamance	North Carolina	B01003_001	166144	NA
37015	Bertie	North Carolina	B01003_001	19081	NA
37063	Durham	North Carolina	B01003_001	317665	NA
37197	Yadkin	North Carolina	B01003_001	37589	NA

Save cleaned data

Save cleaned data

> Tip: `here::here()` will list the RStudio project directory.

```
library(here)

here()
```

```
[1] "C:/Users/jrl/Documents/casestudy_quarto_default_tier_protocol_applied"
```

Or use relative file paths

```
fs::dir_create(here("data/cleaned"))

nc_pop |>
  sf::st_drop_geometry() |>
  janitor::clean_names() |>
  separate(name, into = c("county", "state"), sep = ",") |>
  mutate(county = str_remove(county, " County")) |>
  rename(population = estimate) |>
  write_csv(file = here("data/cleaned/cleaned-ACS_nc-county-populations.csv"))
```

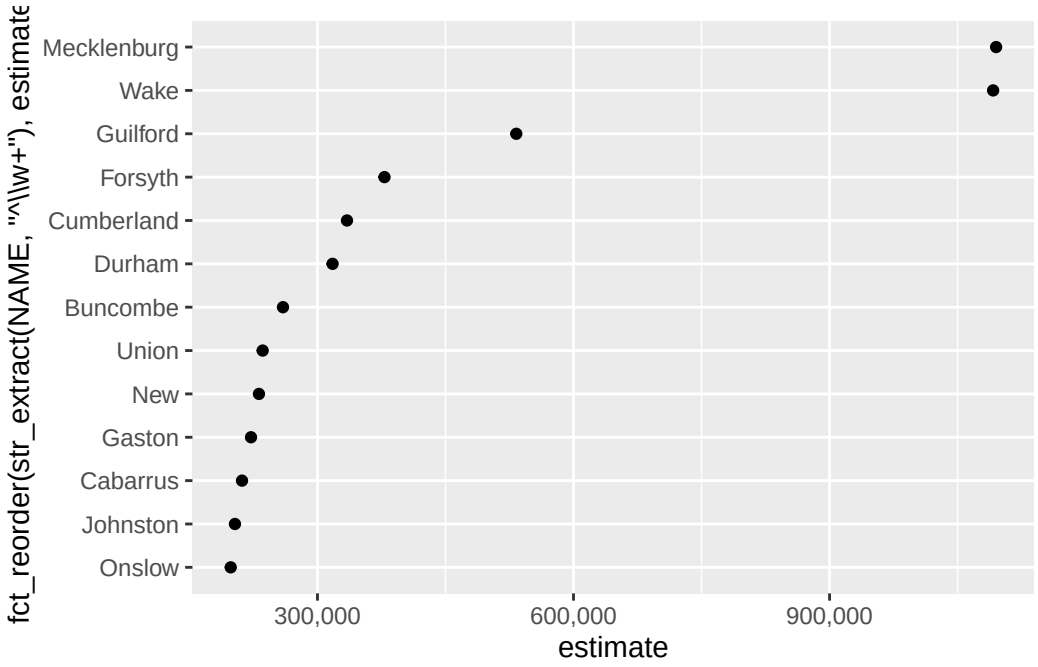
Analysis and visualization

```
nc_pop |>
  sf::st_drop_geometry() |>
  summarise(median_couty_pop = median(estimate),
            mean_county_pop = mean(estimate),
            min_pop = min(estimate),
            max_pop = max(estimate))
```

median_couty_pop	mean_county_pop	min_pop	max_pop
55664	103862.3	3978	1095170

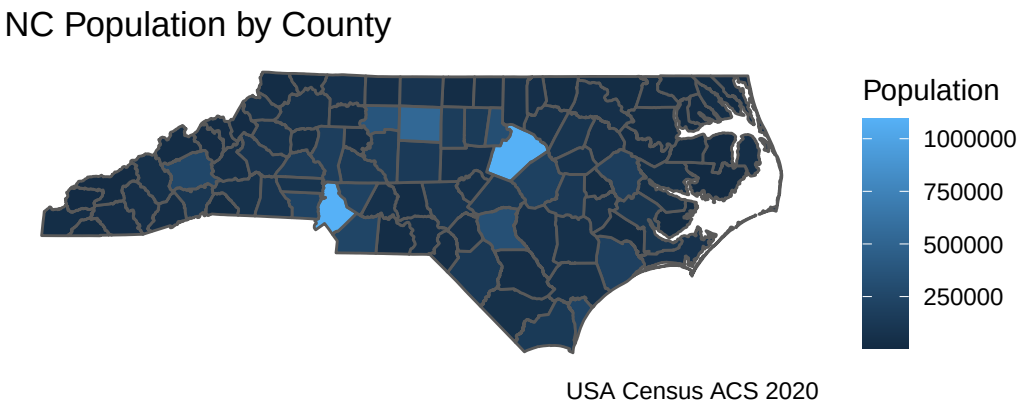
visualization

```
my_viz <- nc_pop %>%
  arrange(desc(estimate)) |>
  slice(1:13) |>
  ggplot(aes(x = estimate, y = fct_reorder(str_extract(NAME, "^\\w+"),
                                           estimate))) +
  geom_point() +
  scale_x_continuous(labels = scales::comma)
my_viz
```



map

```
nc_pop |>
  ggplot(aes(fill = estimate)) +
  geom_sf() +
  coord_sf(datum = NA) +
  theme_minimal() +
  labs(fill = "Population",
       title = "NC Population by County",
       caption = "USA Census ACS 2020")
```



Generate independent outputs

By "independent" I mean manuall, via code, saving outputs to the local file system. This process is not strictly necessary since the quarto compuation notebook includes generated visualizations in the derived reports. But, sometimes we like a belt and suspenders.

Above we used the {here} package to ensure we managed our files relative to the project's root directory on the local file system. You can also use other unix-style relative path constructions such as `..` Using relative file paths in this way accomplishes the same action as {here}. I have included comments using {here} below.

```
# fs::dir_create("../output_secondary/images")
dir_create(here("output_secondary/images"))
# ggsave("../output_secondary/images/top-population_scatter-plot.svg", my_viz)
ggsave(here("output_secondary/images/top-population_scatter-plot.svg"), my_viz)
```

Fin