CHRISTOPHER PRENER, PH.D.
SPRING, 2017

WEEK 12
LECTURE 14

INTRO TO GIS$_C$

WORKING WITH DATA (2)

AGENDA

1. Follow-up
2. GISc & Public Policy
3. Open-Sourcing Spatial Data
4. Shapefiles & Stata
5. Table Joins in Stata
6. Plots in Stata

1 FOLLOW-UP

2 GISc & PUBLIC POLICY

3 OPEN-SOURCING SPATIAL DATA

SHAPEFILES ARE...

- ...ubiquitous
- ...largely open standard but not fully open
- ...hard to work with (they require special software)
- ...tricky to transport

JSON

- Open source, structured data format that is both human and machine readable. Sometimes used for sharing data.

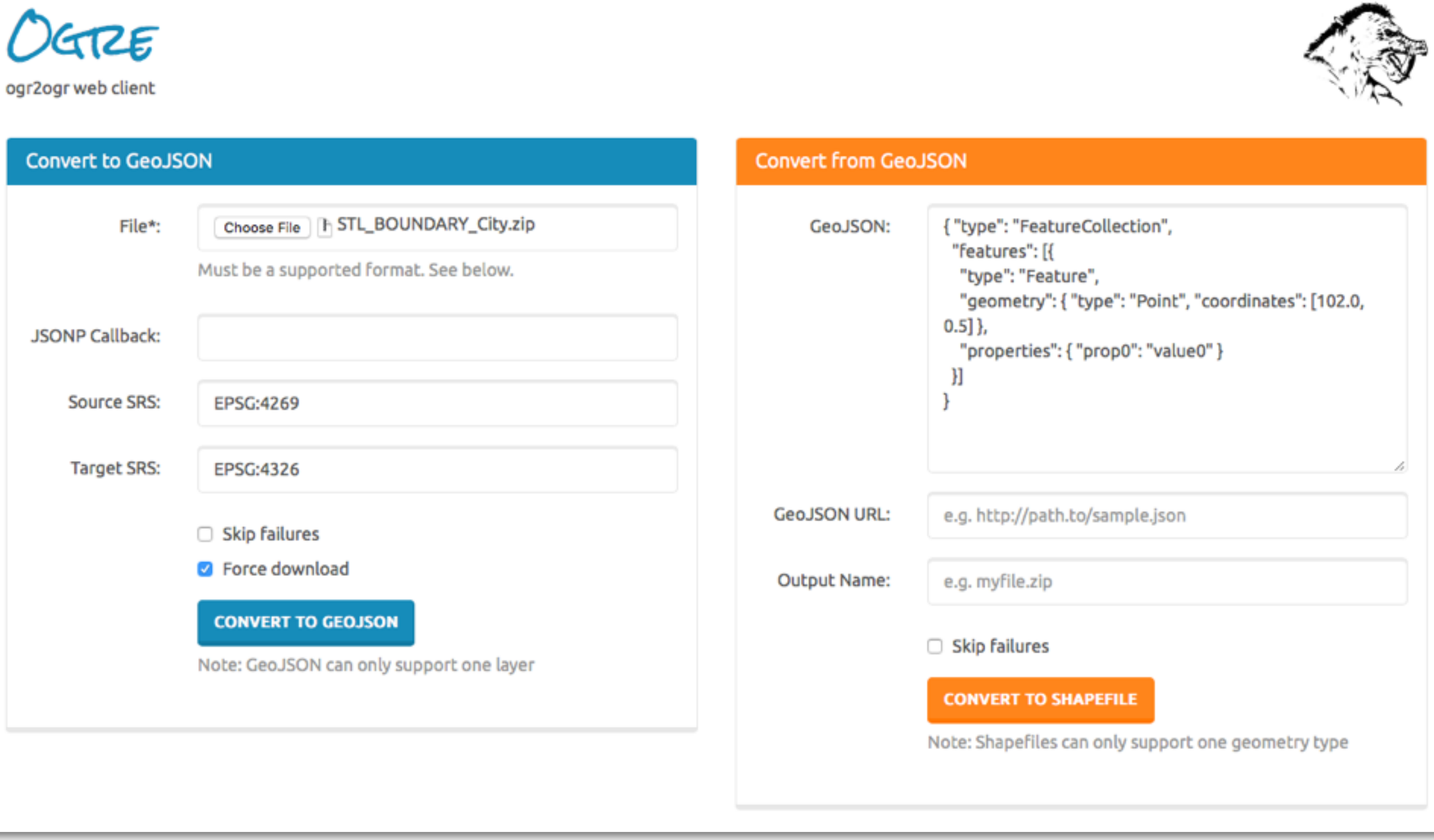
```
{
  "countyName": “St. Louis City”,
  "countyAbbrev": "STL",
  "stateFIPS": 29,
  "countyFIPS": 510
}
```

GEOJSON

- Open source, structured data format that is both human and machine readable **for spatial data.**
- GitHub will project and display these data in your browser!

```
"features": [
    {
      "type": "Feature",
      "properties": {
            "OBJECTID": 1
            "countyName": “St. Louis City”,
            "countyAbbrev": "STL",
            "stateFIPS": 29,
            "countyFIPS": 510
            }
     },
    "geometry": {
            "type": "Polygon",
            "coordinates": [
              [
                [-90.18220619786821,
                 38.77413361698079
                ],
                [-90.1822307573389,
                 38.76859806000061
                ]
              ]
    }
]
```

EXPORTING TO GEOJSON



FORMATTING GEOJSON WITH MAPBOX

STL_BOUNDARY_City.geojson


```
{"type":"FeatureCollection","crs":{"type":"name","properties":{"name":"urn:ogc:def:crs:OGC:1.3:CRS84"}},"features":[{"type":"Feature","properties":{"LAYER":"--CITY_LIMIT_BOUND_2002--","AREA_SQMI":62.11,"Name":"St. Louis","Shape_Leng":218275.771622,"Shape_Area":1845207811.05},"geometry":{"type":"Polygon","coordinates":[[[-90.25039997037375,38.54261303673594],[-90.25777784934488,38.53200017239113],[-90.25778549040501,38.531989181830895],[-90.26062855943232,38.53314561937703],[-90.26310163216301,38.53404792043586],[-90.26374739552416,38.53381697170136],[-90.26443496844155,38.53369895636384],[-90.26543433362905,38.53368866347928],[-90.26705523347422,38.534275062835626],[-90.26817209224178,38.5358051404418],[-90.2685195395497,38.53695651908491],[-90.26907778342093,38.537531610620476],[-90.27033335131908,38.53819138143409],[-90.27104294502718,38.538651978681 32],[-90.27139256132226,38.53928177817347],[-90.27138390270296,38.54010748380952],[-90.2705216571265,38.54162431418637],[-90.27171696957703,38.544541097 1007],[-90.2715830798244,38.544830570324656],[-90.27179615379819,38.545737317771 76],[-90.27218925323444,38.54675712269582],[-90.27225854602999,38.54814140050097],[-90.27230800369968,38.548312827831815],[-90.2723458445704,38.54838725659116],[-90.27410123838251,38.54935404330736],[-90.27791765628682,38.55052068861653],[-90.27903674308176,38.55129333601412],[-90.27933579487264,38.55150461680074],[-90.28385534395238,38.55393764391664],[-90.28719273386477,38.55553370003238],[-90.28977424793635,38.557043472330285],[-90.29239884791451,38.55896009866003],[-90.29460625198516,38.56055849442517],[-90.29694475563973,38.562279696535896],[-90.2971035206985,38.56242548005447],[-90.31360003229126,38.57874722099816],[-90.3162311 14937842,38.58461909401071],[-90.31631986744169,38.58481621292476],[-90.32051233968697,38.59414565179108],[-90.31090640978016,38.61827866874186],[-90.31103118069258,38.61836104167966],[-90.30870714882124,38.63021624407907],[-90.30782134949793,38.632439895557944],[-90.30645464205372,38.6334666376111],[-90.30481236056353,38.64244435707591],[-90.30360893560423,38.64495054487312],[-90.30182443179869,38.655660549456535],[-90.25175051022369,38.71885655967748],[-90.2280525154036,38.72815710573767],[-90.22430557094117,38.72912867486645],[-90.2217704007634,38.730123257265596],[-90.21977892695935,38.73230254962969],[-90.21835237299283,38.7333894929709],[-90.21798185848527,38.73355909110199],[-90.21264898166272,38.736234181901565],[-90.20795420263867,38.73971407164361],[-90.20413669283406,38.743177154674626],[-90.20199472460575,38.74508493321325],[-90.19732636005598,38.749291155380334],[-90.19546135001993,38.751532090427744],[-90.19198859316775,38.754775285 72729],[-90.18925852215104,38.757793868278995],[-90.18831770925426,38.758735251 49595],[-90.18805797514118,38.75906856506351],[-90.18628475757968,38.761369816295 81],[-90.18506240740638,38.764829342533865],[-90.18442902409434,38.768177115351136],[-90.18387091132946,38.7706231240458],[-90.1841940006991,38.77433955000072],[-90.1751202693701,38.77339924313035],[-90.16644780364939,38.77274185198855],[-90.16630272872025,38.77270511849884],[-90.16640899671849,38.77264900055753],[-90.
```

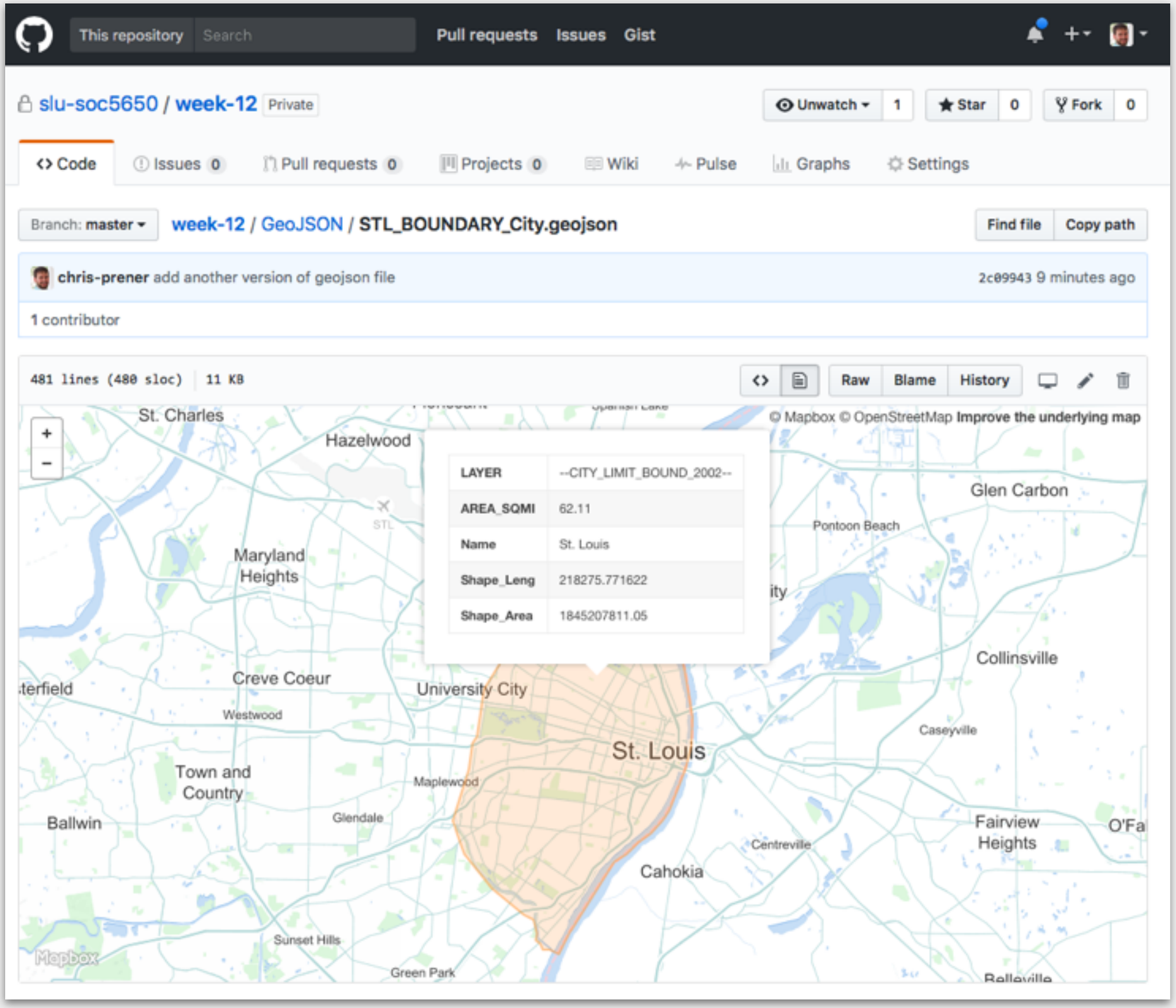
FORMATTING GEOJSON WITH MAPBOX

GeoJSON Hint

↔ format


```
{
  "type": "FeatureCollection",
  "crs": {
    "type": "name",
    "properties": {
      "name": "urn:ogc:def:crs:OGC:1.3:CRS84"
    }
  },
  "features": [
    {
      "type": "Feature",
      "properties": {
        "LAYER": "--CITY_LIMIT_BOUND_2002--",
        "AREA_SQMI": 62.11,
        "Name": "St. Louis",
        "Shape_Leng": 218275.771622,
        "Shape_Area": 1845207811.05
      },
      "geometry": {
        "type": "Polygon",
        "coordinates": [
          [
            [
              -90.25039997037375,
              38.54261303673594
            ],
            [
              -90.25777784934488,
              38.53200017239113
            ],
            [
              -90.25778549040501,
              38.531989181830895
            ],
            [
              -90.260628559432,
              38.53314561937703
            ],
            [
              -90.26310163216301,
              38.53404792043586
            ],
            [
              -90.26374739552416,
              38.53381697170136
            ],
            [
              -90.26443496844155,
              38.53369895636384
            ],
            [
              -90.26543433362905,
              38.53368866347928
            ],
            [
              -90.26705523347422,
              38.534275062835626
            ],
```

ARCHIVING GEOJSON WITH GITHUB



4 SHAPEFILES & STATA

SHAPEFILES ARE. . .

- … difficult to edit in a reproducible fashion
- … often full of data we do not need
- … often inconsistently organized

SHP2DTA PACKAGE

- Offers a set of tools for importing shapefile data into Stata
- Converts shapefiles into two Stata datasets - one for the attribute table and one for the coordinate data.

```
ssc install shp2dta



shp2dta using shapefile, database(dataFile)
coordinates(coordFile) genid(id)
```

EXAMPLE

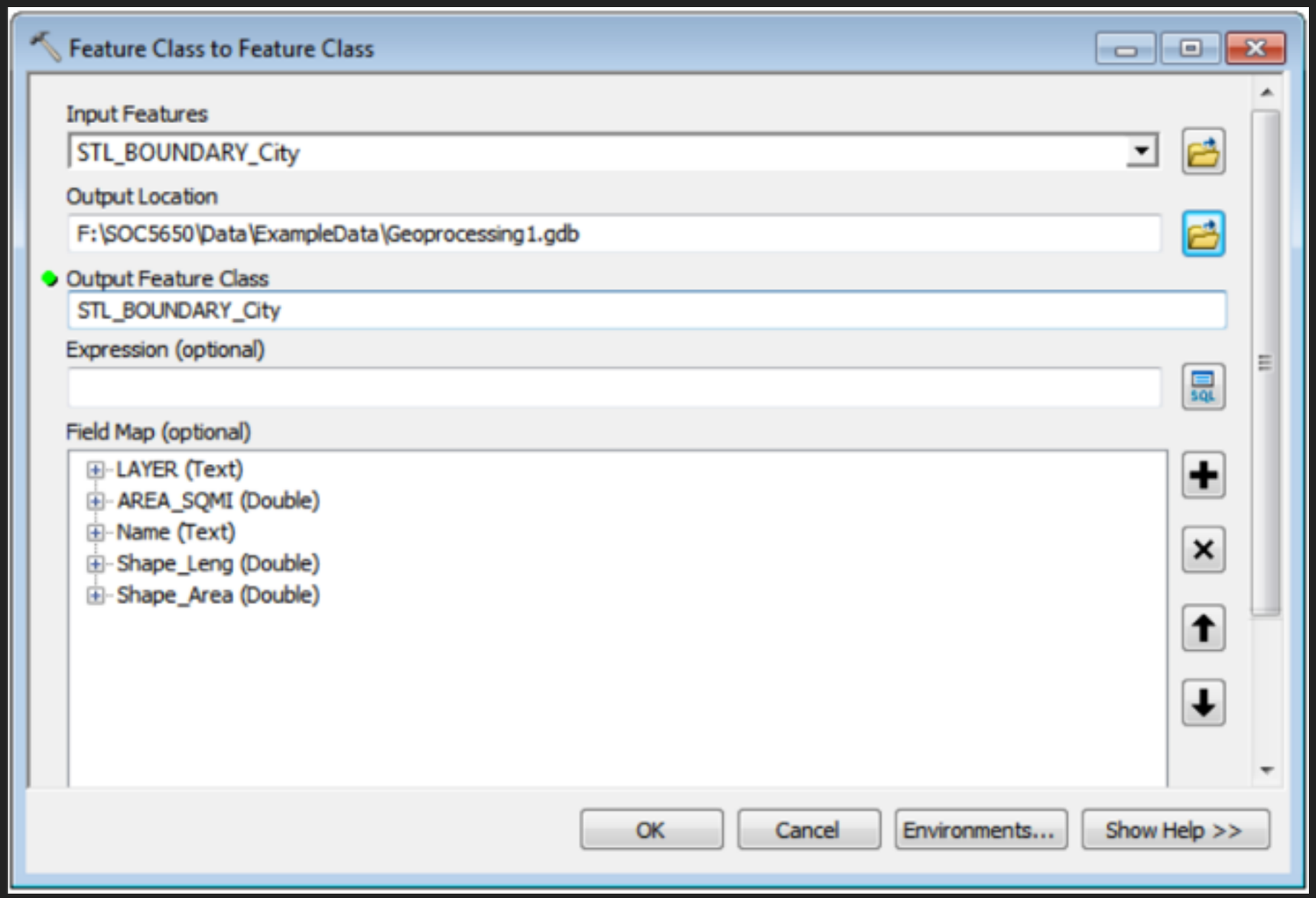
```
. shp2dta using STL_BOUNDARY_City, database(cityData) coordinates(cityCoord) genid(id)
type: 5


. use cityData.dta
```

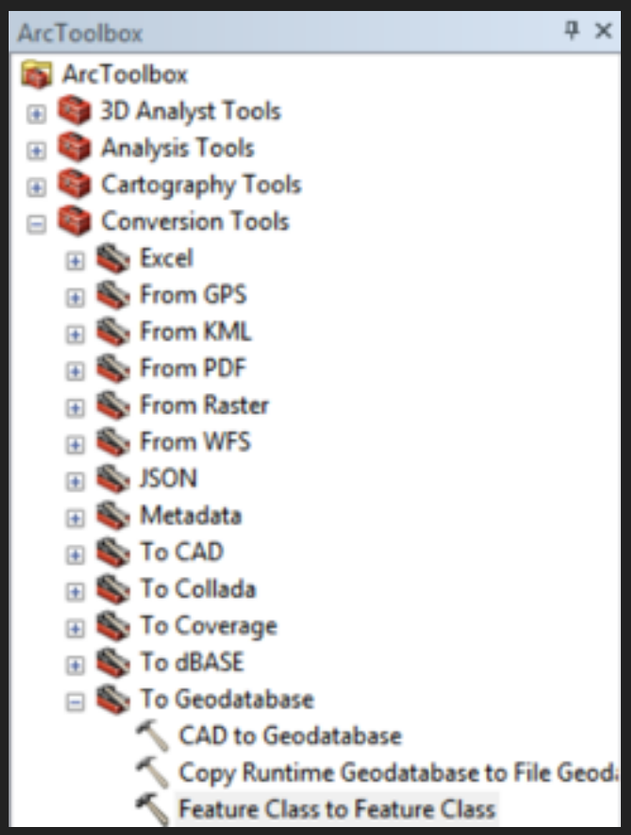
```
. describe

Contains data from cityData.dta
  obs:             1
 vars:             6                          4 Apr 2017 13:40
 size:            47
-------------------------------------------------------------------------------
              storage   display    value
variable name   type    format     label      variable label
-------------------------------------------------------------------------------
LAYER           str25   %25s                  LAYER
AREA_SQMI       float   %9.0g                 AREA_SQMI
Name            str9    %9s                   Name
Shape_Leng      float   %9.0g                 Shape_Leng
Shape_Area      float   %9.0g                 Shape_Area
id              byte    %12.0g
-------------------------------------------------------------------------------
Sorted by: id
```

USING SHP2DTA

1. Import the shapefile into Stata
2. Using data cleaning commands to alter as needed (create new variables, drop variables, clean variables)
3. Save data and export as a csv and/or Excel file
4. Export a “clean” version of the shapefile to a new feature class that contains only the id variable
5. Use a table join to combine them in ArcGIS

EXPORTING "CLEAN" FEATURE CLASSES





5 TABLE JOINS IN STATA

CONCEPTUAL REVIEW OF TABLE JOINS

a	b
1	St. Louis City
2	St. Louis
3	Franklin
4	Jefferson

c	d
1	318,416
2	1,003,362
3	101,816
4	221,396

CONCEPTUAL REVIEW OF TABLE JOINS

a	b	d
1	St. Louis City	318,416
2	St. Louis	1,003,362
3	Franklin	101,816
4	Jefferson	221,396

TABLE JOINS IN ARCGIS ARE. . .

- … like shapefiles - difficult to do in a reproducible fashion
- … prone to conflicts over data types that can be frustrating to manage

SOME DEFINITIONS

- **Master Dataset** - dataset currently in Stata's memory
- **Using Dataset** - second Stata dataset that contains additional data
- **Key -** identification variable that is named the same in both datasets *and* stored in the same format

EXECUTING TABLE JOINS IN STATA

```
merge 1:1 keyVar using usingData.dta
```

```
. use STL_HEALTH_Lead.dta
```

```
. merge 1:1 tractce using tractData.dta
    Result                           # of obs.
    -----------------------------------------
    not matched                             0
    matched                               106  (_merge==3)
    -----------------------------------------
```

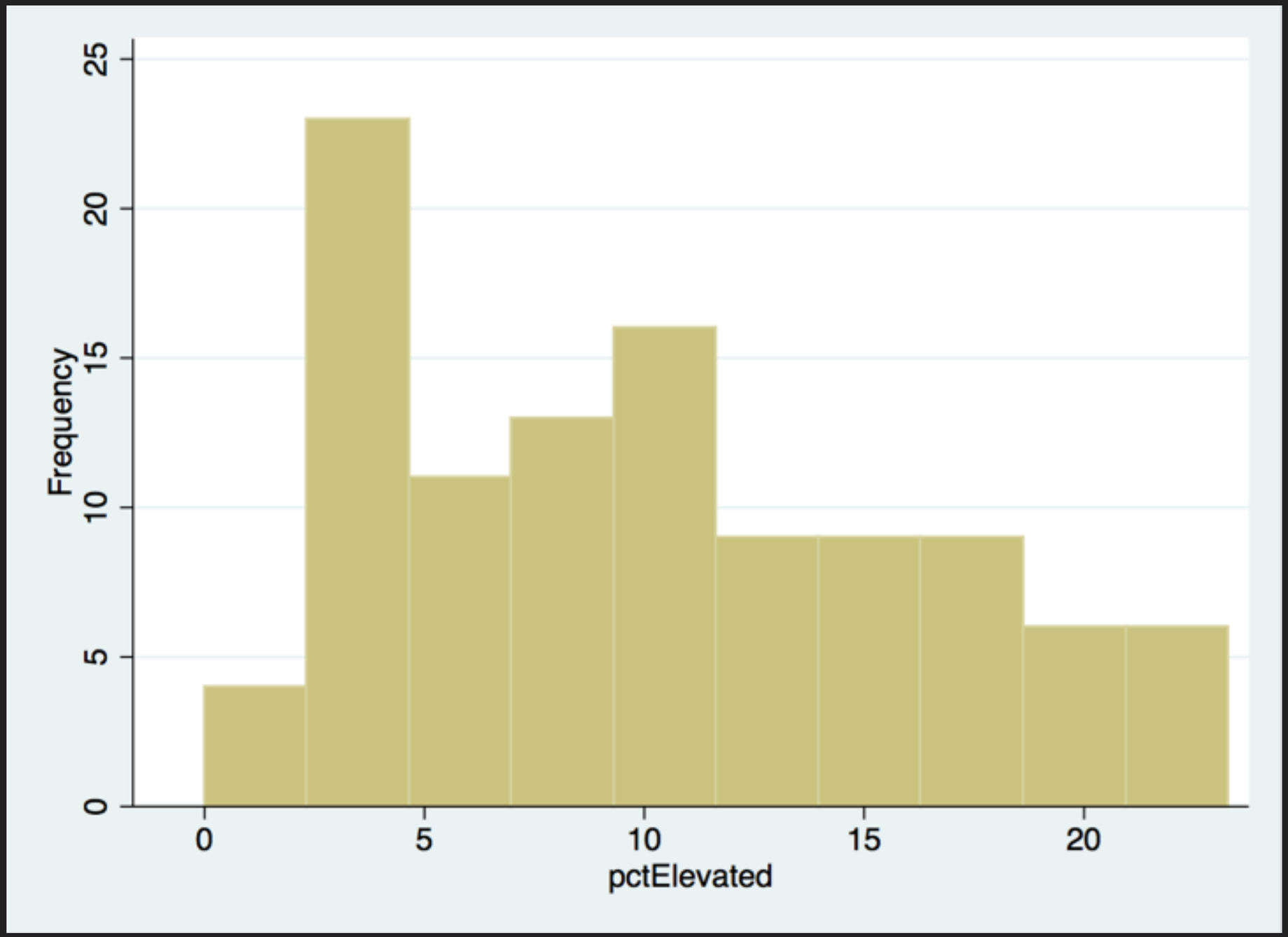
6 PLOTS IN STATA

HISTOGRAMS

```
histogram varname, frequency bin(binNumber)
```

```
. histogram pctelevated, frequency
(bin=10, start=0, width=2.3280001)
```

HISTOGRAMS

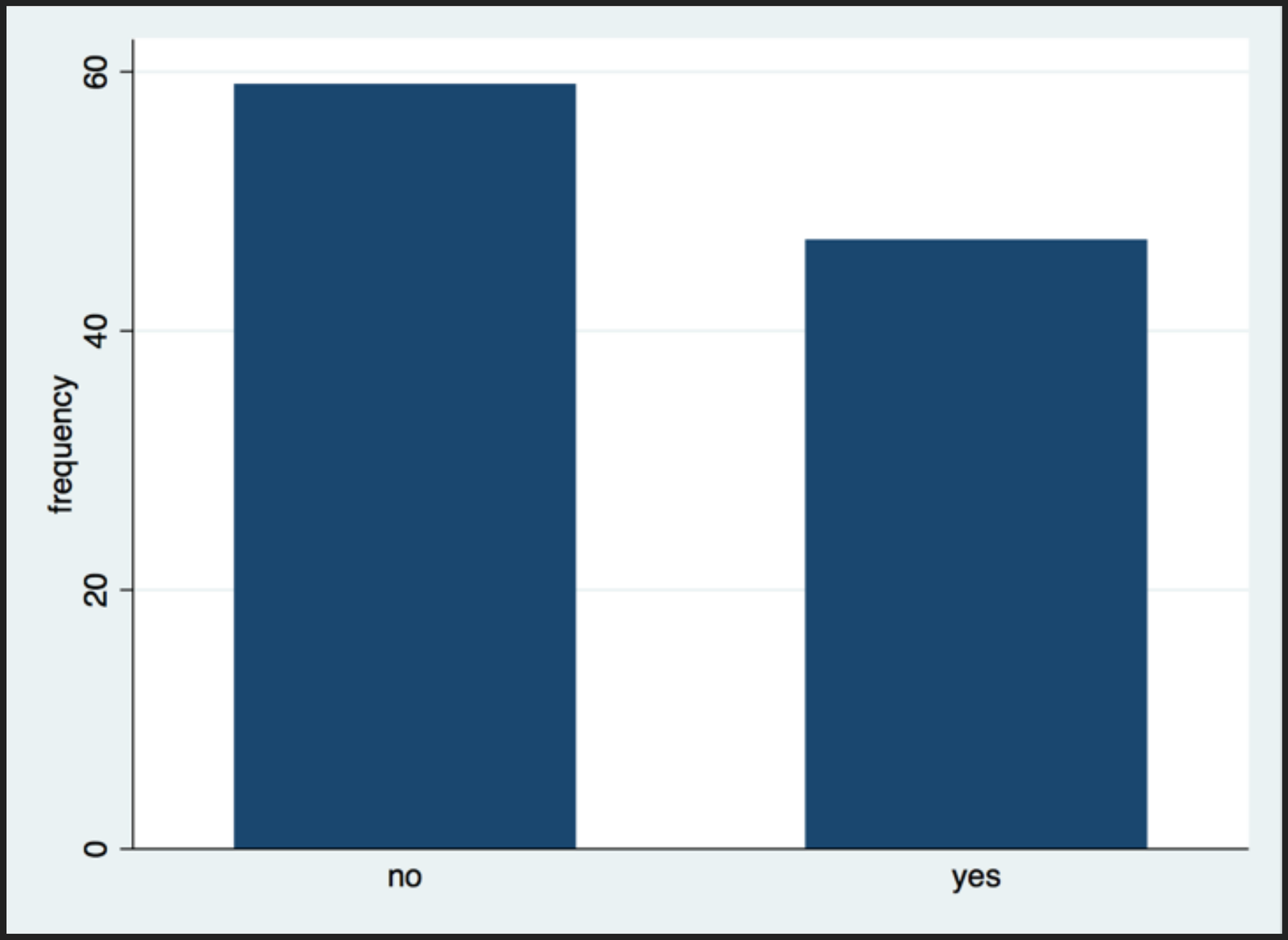


VERTICAL BAR CHARTS

```
graph bar (count), over(varname)
```

```
. graph bar (count), over(pctelevated)
```

VERTICAL BAR CHARTS

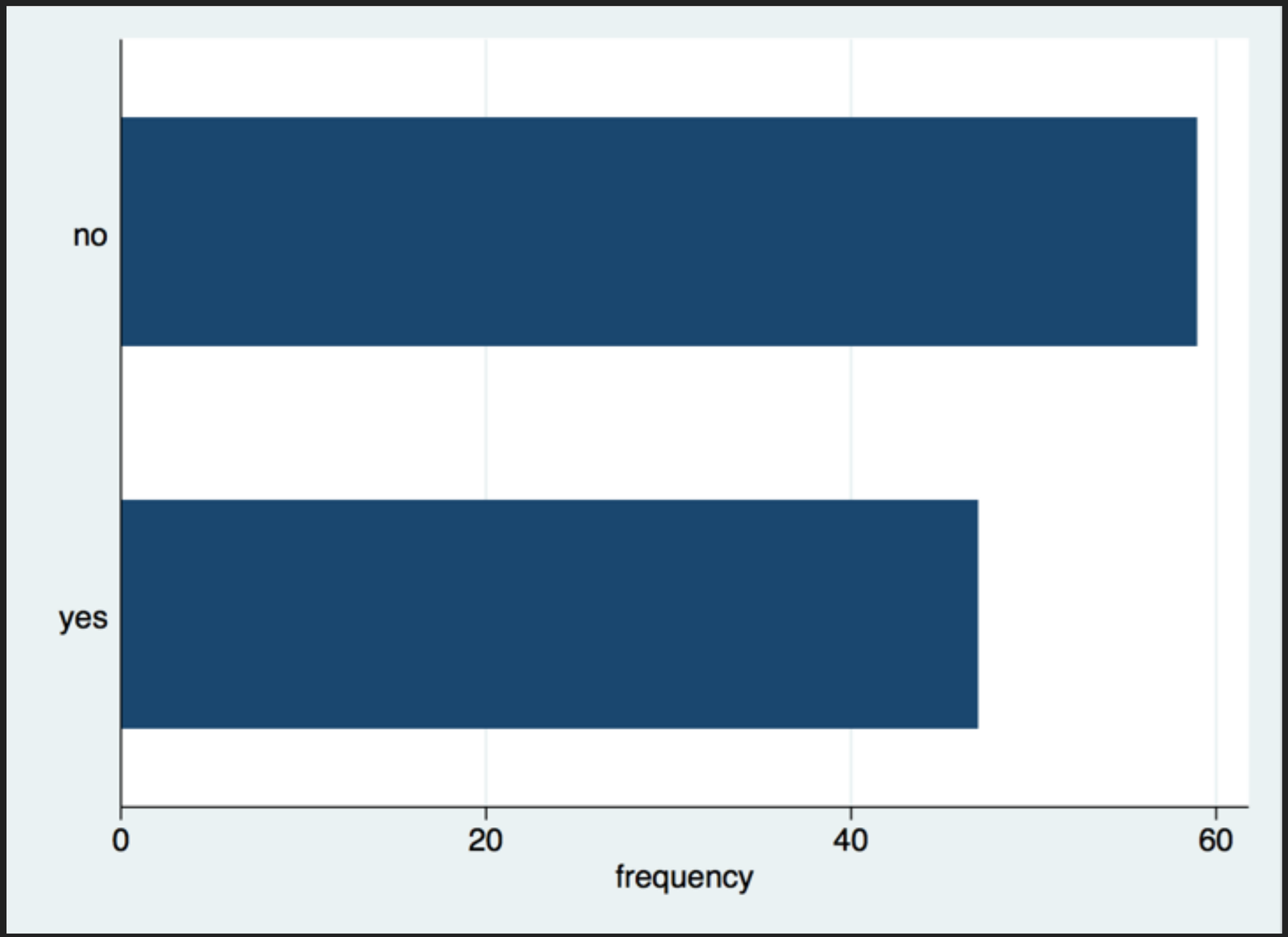


HORIZONTAL BAR CHARTS

```
graph hbar (count), over(varname)
```

```
. graph hbar (count), over(pctelevated)
```

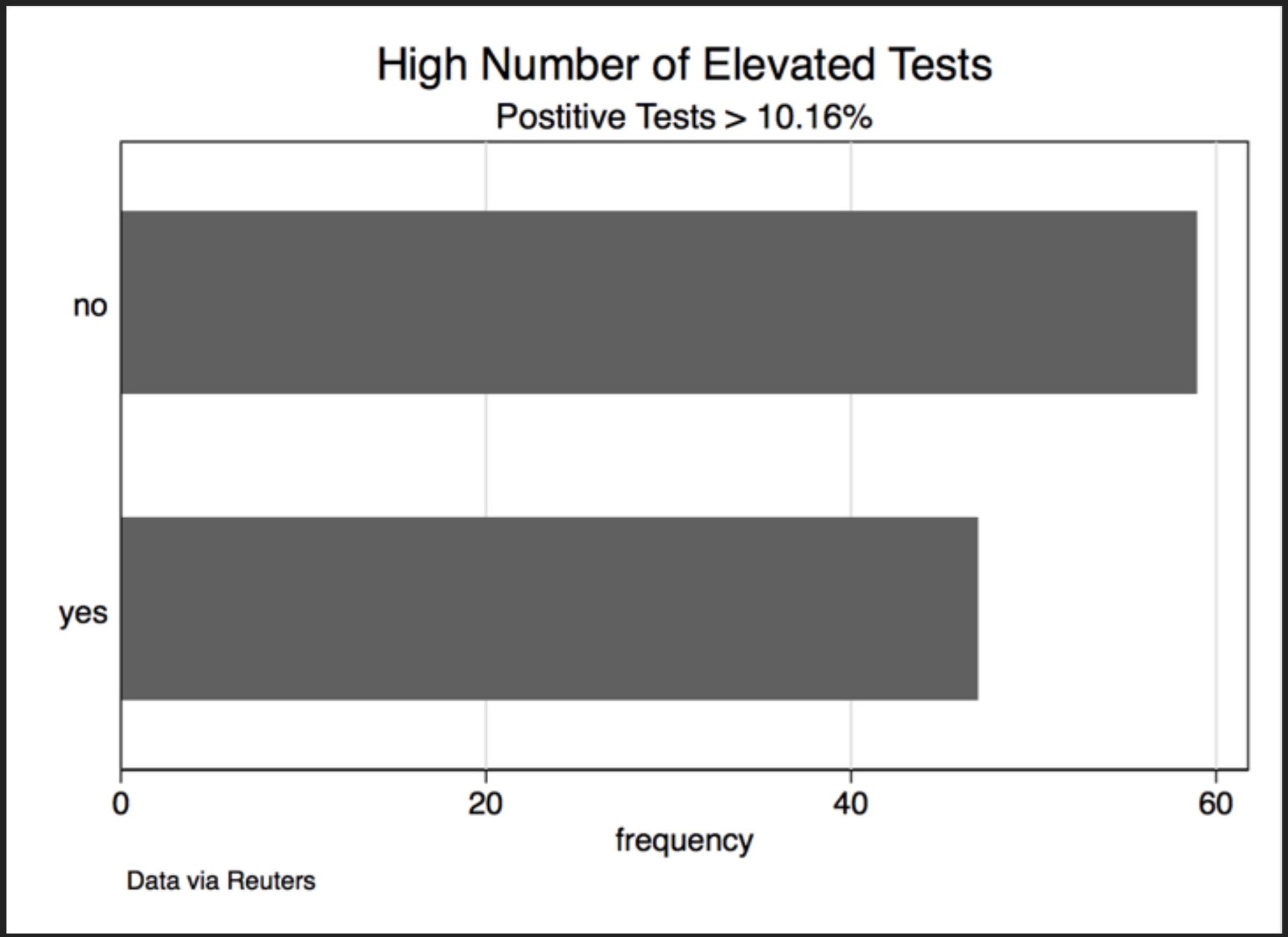
HORIZONTAL BAR CHARTS



GRAPH OPTIONS

```
title(“title text”)

subtitle(“subtitle text”)

note(“note text”)

scheme(schemeName)


. graph hbar (count), over(varname) title(“title text”) ///
    subtitle(“subtitle text”) note(“note text”) scheme(scheme)
```

GRAPH OPTIONS



EXPORTING GRAPHS

```
graph export “filename.png”, width(val) height(val) replace
```

```
. graph export “leadHistogram.png”, width(800) height(600) replace
```