```
//This script makes the summary statistics and does the regression for the
paper.
//Copyleft Ed Egan, 2022

//Open the data
insheet using "MasterCBSALayersv20-7.txt"
set more off
log using V20-CBSA-7-0.log, replace

//First build the id
egen cbsaid=group(cbsa)

//Remove MICRO-SAs (Lebanon, NH-VT, cbsaid=40)
drop if cbsa=="Lebanon, NH-VT"

//Define Variables
egen cbsalayerid=group(cbsa layer)
sort cbsalayerid year
xtset cbsalayerid year
gen growthinv20=growthinv*inflator20
gen growthinvcluster20=growthinvcluster*inflator20
gen growthinv20f = f.growthinv20
gen growthinv20l=ln(1+growthinv20)
gen growthinv20lf = f.growthinv20l
gen nosinglemulti=nosingleton+nomultiton
gen totsinglemulticount=totmultitoncount+nosingleton
gen vcr2eligible=0
replace vcr2eligible=1 if lowesthighestflag==1 & year>=2002
gen fracinhulls = tothullcount/numstartups
gen growthinv20diff=f.growthinv20-growthinv20
gen avghulldistkm=avghulldisthm/10
gen avghulldensityl=ln(avghulldensity)
gen tothulldensityl=ln(tothulldensity)
gen avghullareal=ln(avghullarea)
gen numstartupsl=ln(numstartups)
gen exclude=0
replace exclude=1 if cbsaid==53 | cbsaid==68
gen top3density=0
replace top3density=1 if cbsaid==19| cbsaid==54 | cbsaid==74
gen top5density=0
replace top5density=1 if cbsaid==19 | cbsaid==54 | cbsaid==74 | cbsaid==10 |
cbsaid==26
gen top5numstarts=0
replace top5numstarts=1 if cbsaid==10 | cbsaid==42 | cbsaid==54 | cbsaid==74 |
cbsaid==75

//Descriptive stats for the table (note that exclude=0 throws out 53: New
Orleans-Metairie, LA and 68: Riverside-San Bernardino-Ontario, CA, which have
single massive areas with any max r2 method)
estpost tabstat layerindex numclusters fracinhulls tothullcount avghullarea
tothulldensity avghulldistkm numstartups growthinv20 if layer==elbowlayer,
stats(p50 mean sd) columns(statistics)
esttab using table1.csv, replace cell((p50 mean sd)) nonumber nomtitle
estpost tabstat layerindex numclusters fracinhulls tothullcount avghullarea
tothulldensity avghulldistkm numstartups growthinv20 if layer==cubicmaxlayer,
stats(p50 mean sd) columns(statistics)
esttab using table1.csv, append cell((p50 mean sd)) nonumber nomtitle

//Regressions for table (note that this is for the CubicMaxLayers...)
reg growthinv20l tothulldensityl if layer==cubicmaxlayer, cluster(cbsaid)
outreg2 using regressionv2, excel replace
reg growthinv20l tothulldensityl fracinhulls tothullcount avghullareal
avghulldisthm if layer==cubicmaxlayer, cluster(cbsaid)
outreg2 using regressionv2, excel append
```

```
reg growthinv20l tothulldensityl fracinhulls tothullcount avghullareal
avghulldisthm numstartupsl if layer==cubicmaxlayer, cluster(cbsaid)
outreg2 using regressionv2, excel append
reg growthinv20l tothulldensityl fracinhulls tothullcount avghullareal
avghulldisthm numstartupsl i.year if layer==cubicmaxlayer, cluster(cbsaid)
outreg2 using regressionv2, excel append
```