R Code for Illustration

Prof. Jude C. Hays

3/3/2021

Making a map with ggplot, sf, and tigris

```
us <- read_sf("gadm36_USA_shp/gadm36_USA_1.shp")
oxford = read.csv("oxford_data.csv")
head(us)
```

```
## Simple feature collection with 6 features and 10 fields
## geometry type:  MULTIPOLYGON
## dimension:      XY
## bbox:           xmin: -179.1506 ymin: 30.21725 xmax: 179.7734 ymax: 72.6875
## geographic CRS: WGS 84
## # A tibble: 6 x 11
##   GID_0 NAME_0 GID_1 NAME_1 VARNAME_1 NL_NAME_1 TYPE_1 ENGTYPE_1 CC_1  HASC_1
##   <chr> <chr>  <chr> <chr>  <chr>     <chr>     <chr>  <chr>     <chr> <chr>
## 1 USA   Unite~ USA.~ Alaba~ AL|Ala.   <NA>      State  State     <NA>  US.AL
## 2 USA   Unite~ USA.~ Alaska AK|Alaska <NA>      State  State     <NA>  US.AK
## 3 USA   Unite~ USA.~ Arizo~ AZ|Ariz.  <NA>      State  State     <NA>  US.AZ
## 4 USA   Unite~ USA.~ Arkan~ AR|Ark.   <NA>      State  State     <NA>  US.AR
## 5 USA   Unite~ USA.~ Calif~ CA|Calif. <NA>      State  State     <NA>  US.CA
## 6 USA   Unite~ USA.~ Color~ CO|Colo.  <NA>      State  State     <NA>  US.CO
## # ... with 1 more variable: geometry <MULTIPOLYGON [°]>
```

```
head(oxford)
```

```
##   statenm    ben95  rskpovpc  wage95 instcoad   ipcfold teitrend  match state
## 1      AL 204.1363 18.478960 1120.238   37.230 0.8051326     84.0 72.948     1
## 2      AZ 473.6823 14.727540 1210.826   15.994 0.9642222     98.3 61.830     2
## 3      AR 337.3343 20.932880 1162.445   71.626 0.7160101     85.7 74.414     3
## 4      CA 952.1440 11.846200 1263.137   48.228 0.9345878     94.5 50.000     4
## 5      CO 558.8927 10.712840 1180.645   47.456 0.9578660     86.6 51.140     5
## 6      CT 718.2612  6.096505 1066.758   87.896 0.8663776     93.0 50.000     6
##   objectid abrev        name fips        lon      lat shape_leng shape_area id
## 1        2    AL     Alabama    1  -86.82676 32.79354   27.83579  12.870620  2
## 2        5    AZ     Arizona    4 -111.66460 34.29323   23.83830  28.921379  5
## 3        3    AR    Arkansas    5  -92.43920 34.89977   21.80720  13.585590  3
## 4        6    CA  California    6 -119.60820 37.24537   63.38137  41.623650  6
## 5        7    CO    Colorado    8 -105.54780 38.99855   22.02048  28.039440  7
## 6        8    CT Connecticut    9  -72.72623 41.62196   13.60293   1.390348  8
```

```
data <- geo_join(
  spatial_data=us,
  data_frame = oxford,
  by_sp = "NAME_1",
```

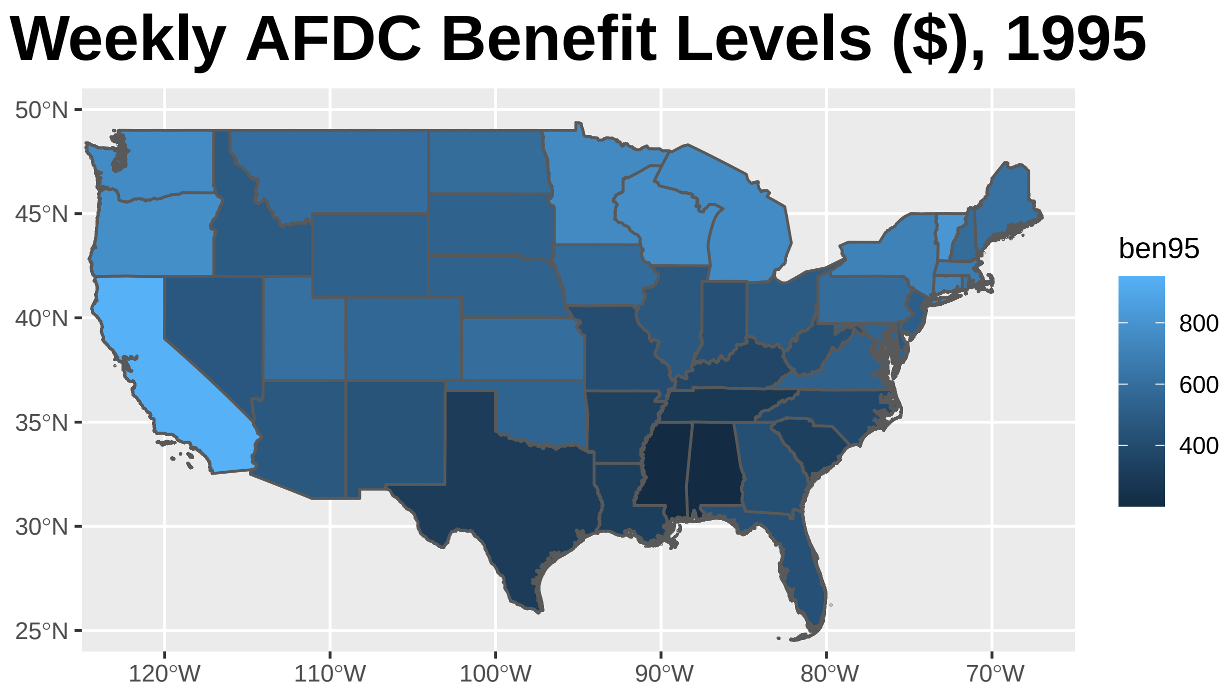
```r
  by_df = "name",
  how = "inner"
)

head(data)
```

```
## Simple feature collection with 6 features and 27 fields
## geometry type:  MULTIPOLYGON
## dimension:      XY
## bbox:           xmin: -124.4156 ymin: 30.21725 xmax: -71.78824 ymax: 42.04973
## geographic CRS: WGS 84
## # A tibble: 6 x 28
##   GID_0 NAME_0 GID_1 NAME_1 VARNAME_1 NL_NAME_1 TYPE_1 ENGTYPE_1 CC_1  HASC_1
##   <chr> <chr>  <chr> <chr>  <chr>     <chr>     <chr>  <chr>     <chr> <chr>
## 1 USA   Unite~ USA.~ Alaba~ AL|Ala.   <NA>      State  State     <NA>  US.AL
## 2 USA   Unite~ USA.~ Arizo~ AZ|Ariz.  <NA>      State  State     <NA>  US.AZ
## 3 USA   Unite~ USA.~ Arkan~ AR|Ark.   <NA>      State  State     <NA>  US.AR
## 4 USA   Unite~ USA.~ Calif~ CA|Calif. <NA>      State  State     <NA>  US.CA
## 5 USA   Unite~ USA.~ Color~ CO|Colo.  <NA>      State  State     <NA>  US.CO
## 6 USA   Unite~ USA.~ Conne~ CT|Conn.  <NA>      State  State     <NA>  US.CT
## # ... with 18 more variables: geometry <MULTIPOLYGON [°]>, statenm <chr>,
## #   ben95 <dbl>, rskpovpc <dbl>, wage95 <dbl>, instcoad <dbl>, ipcfold <dbl>,
## #   teitrend <dbl>, match <dbl>, state <int>, objectid <int>, abrev <chr>,
## #   fips <int>, lon <dbl>, lat <dbl>, shape_leng <dbl>, shape_area <dbl>,
## #   id <int>
```

```r
usplot <- ggplot(data, aes(fill=ben95)) +
  geom_sf() +
   coord_sf(xlim = c(-125, -65), ylim = c(24, 51), expand = FALSE)+
  ggtitle("Weekly AFDC Benefit Levels ($), 1995")+
  theme(plot.title=element_text(hjust=.5)) +
  theme(plot.title=element_text(size=24,face="bold"))

usplot
```



Do some analysis

```
attach(oxford)
X = as.matrix(oxford[,4:9])


weights = read.csv("oxford_w.csv")
weights = as.matrix(weights)


SLX = weights %*% X

weights = mat2listw(weights, row.names=NULL)
weights = nb2listw(weights$neighbours, style="W")
#row-standardize the matrix

#Non-spatial regression

oxford.ols = lm(ben95 ~ rskpovpc + wage95 + instcoad + ipcfold + teitrend + match)

#LM and Robust LM tests

lmtests <- lm.LMtests(oxford.ols, weights, test = c("LMerr","LMlag","RLMerr","RLMlag"))
summary(lmtests)
```

```
##  Lagrange multiplier diagnostics for spatial dependence
## data:
## model: lm(formula = ben95 ~ rskpovpc + wage95 + instcoad + ipcfold +
## teitrend + match)
## weights: weights
##
##        statistic parameter   p.value
## LMerr    5.84473         1 0.0156238 *
## LMlag   11.60584         1 0.0006574 ***
## RLMerr   0.71624         1 0.3973796
## RLMlag   6.47736         1 0.0109257 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#Spatial AutoRegressive (SAR) lag model

lag = lagsarlm(ben95 ~ rskpovpc +wage95 + instcoad + ipcfold + teitrend +match, data=oxford, weights ,m
summary(lag)
```

```
##
## Call:lagsarlm(formula = ben95 ~ rskpovpc + wage95 + instcoad + ipcfold +
##     teitrend + match, data = oxford, listw = weights, method = "eigen",
##     zero.policy = TRUE, tol.solve = 1e-11)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -146.1977  -71.7770   -5.0836   49.8639  344.8755
##
## Type: lag
## Coefficients: (asymptotic standard errors)
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept) -155.41168  429.35523 -0.3620  0.71738
## rskpovpc       3.61445    8.91974  0.4052  0.68532
## wage95        -0.02553    0.18232 -0.1400  0.88864
## instcoad       1.43251    0.80614  1.7770  0.07557
## ipcfold      445.40931  226.99909  1.9622  0.04974
## teitrend       2.42686    1.26262  1.9221  0.05460
## match         -5.38937    3.90333 -1.3807  0.16737
##
## Rho: 0.53475, LR test value: 12.354, p-value: 0.00044009
## Asymptotic standard error: 0.12196
##     z-value: 4.3845, p-value: 1.1627e-05
## Wald statistic: 19.224, p-value: 1.1627e-05
##
## Log likelihood: -287.4067 for lag model
## ML residual variance (sigma squared): 8573.5, (sigma: 92.593)
## Number of observations: 48
## Number of parameters estimated: 9
## AIC: 592.81, (AIC for lm: 603.17)
## LM test for residual autocorrelation
## test value: 1.3773, p-value: 0.24056
```

```
# the default tolerance level (of detecting linear independence of the matrix you try to invert) is , 1

moran.test(resid(oxford.ols), weights)
```

```
##
##  Moran I test under randomisation
##
## data:  resid(oxford.ols)
## weights: weights
##
## Moran I statistic standard deviate = 2.8002, p-value = 0.002554
## alternative hypothesis: greater
## sample estimates:
## Moran I statistic       Expectation          Variance
##       0.246473800      -0.021276596       0.009143082
```

```
moran.test(resid(lag), weights)
```

```
##
##  Moran I test under randomisation
##
## data:  resid(lag)
## weights: weights
##
## Moran I statistic standard deviate = -0.47255, p-value = 0.6817
## alternative hypothesis: greater
## sample estimates:
## Moran I statistic       Expectation          Variance
##      -0.065960243      -0.021276596       0.008941381
```

```
# you can extract the log-likelihood value from the regression
logLik(lag)
```

```
## 'log Lik.' -287.4067 (df=9)
```

```
logLik(oxford.ols)
```

```
## 'log Lik.' -293.5836 (df=8)
```

```
# one way to evaluate the "meaningfulness" of adding the spatial lag ("the others' effects") is to comp
# compute the likelihood ratio.
# "lagsarlm" computes these statistics, but there are ways to get them by hand.
lr1 <- 2*(logLik(lag)-logLik(oxford.ols))

# there is also a canned command for computing the likelihood ratio
lr2 <- LR.sarlm(lag, oxford.ols)

# confirm that the likelihood ratios computed in these two ways are the same
c(lr1[1], lr2[[1]])
```

```
##                   Likelihood ratio
##          12.35379         12.35379
```

```
# similarly, wald test
Wald1.sarlm(lag)
```

```
##
##  Wald diagnostics for spatial dependence
##
## data:
## Wald statistic = 19.224, df = 1, p-value = 1.163e-05
## sample estimates:
```

```
##       rho
## 0.5347474
```

```r
#Spatial error model#
error = errorsarlm(ben95 ~ rskpovpc  +wage95+ instcoad + ipcfold  + teitrend+match, data=oxford, weight
summary(error)
```

```
##
## Call:errorsarlm(formula = ben95 ~ rskpovpc + wage95 + instcoad + ipcfold +
##     teitrend + match, data = oxford, listw = weights, method = "eigen",
##     tol.solve = 1e-20)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -143.691  -56.049  -12.929   38.018  364.777
##
## Type: error
## Coefficients: (asymptotic standard errors)
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) 681.90010  472.06382  1.4445  0.14860
## rskpovpc      3.35826   10.05688  0.3339  0.73843
## wage95       -0.34561    0.24292 -1.4227  0.15481
## instcoad      1.69858    0.82069  2.0697  0.03848
## ipcfold     260.97438  238.08142  1.0962  0.27301
## teitrend      2.92563    1.21123  2.4154  0.01572
## match        -6.90918    4.09402 -1.6876  0.09148
##
## Lambda: 0.57089, LR test value: 8.4244, p-value: 0.0037022
## Asymptotic standard error: 0.13366
##     z-value: 4.2713, p-value: 1.943e-05
## Wald statistic: 18.244, p-value: 1.943e-05
##
## Log likelihood: -289.3714 for error model
## ML residual variance (sigma squared): 9181.2, (sigma: 95.819)
## Number of observations: 48
## Number of parameters estimated: 9
## AIC: 596.74, (AIC for lm: 603.17)
```

```r
#SLX#
lagx= lm(ben95 ~ rskpovpc + wage95 + instcoad + ipcfold + teitrend + match + SLX[,1] )
summary(lagx)
```

```
##
## Call:
## lm(formula = ben95 ~ rskpovpc + wage95 + instcoad + ipcfold +
##     teitrend + match + SLX[, 1])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -176.44  -76.68   -3.78   59.94  360.38
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 206.26805  535.41517   0.385   0.7021
## rskpovpc     -2.06653   11.54523  -0.179   0.8588
## wage95       -0.15053    0.22487  -0.669   0.5071
```

```
## instcoad      1.54465    1.01738   1.518   0.1368  
## ipcfold     617.94961  287.13168   2.152   0.0375 *
## teitrend      2.97289    1.58876   1.871   0.0686 .
## match        -5.46807    4.99113  -1.096   0.2798  
## SLX[, 1]     -0.01627    0.01128  -1.442   0.1572  
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 117.1 on 40 degrees of freedom
## Multiple R-squared:  0.553,  Adjusted R-squared:  0.4747
## F-statistic: 7.068 on 7 and 40 DF,  p-value: 1.687e-05
```