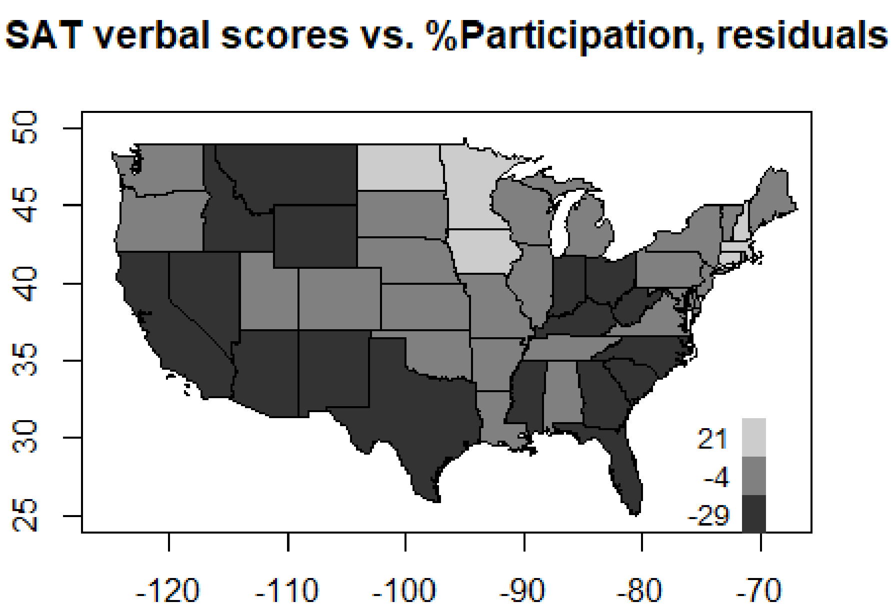
HW4

Steve Harms

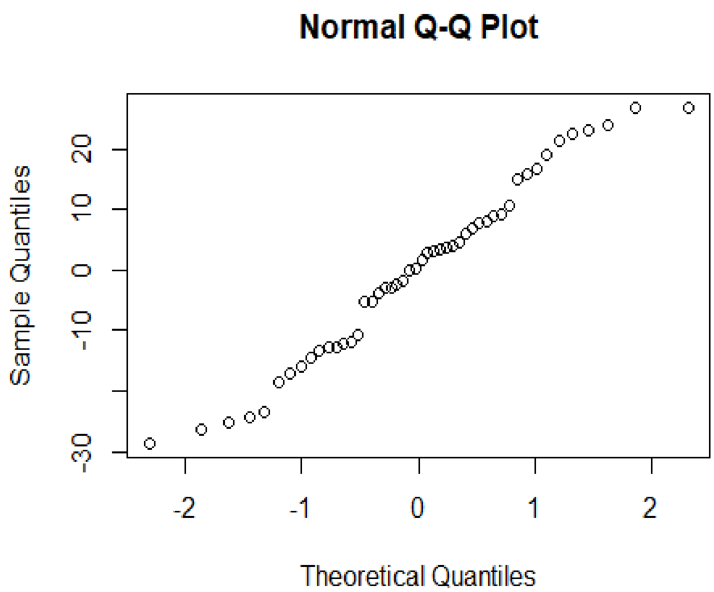
April 4, 2018

a)

```r
satprox <- as.matrix(read.table("sat_proximity.txt", header=F))
sat <- read.table("sat.txt", header=TRUE)
load("US_states.RData")
where.is.state <- pmatch(tolower(US$STATE_NAME), tolower(sat$name))
#simple linear regression model
slr <- lm(vscore~pc, data=sat)
#summarize coefficients
summary(slr)$coefficients
##                Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 572.446896 3.59298945 159.32329 9.751622e-65
## pc           -1.062325 0.07999413 -13.28004 2.384271e-17
#get residuals
residslr <- slr$residuals
#choropeth plot
plot(0, 0, xlim=c(-125, -68), ylim=c(25,50), type="n",
     xlab="", ylab="", main="SAT verbal scores vs. %Participation, residuals")
plot.poly(US, residslr[where.is.state], seq(-29, 27, 25),
          legend.x=c(-71, -70), legend.y=c(25, 30), add=TRUE)
```




```r
#qqplot
qqnorm(residslr)
```

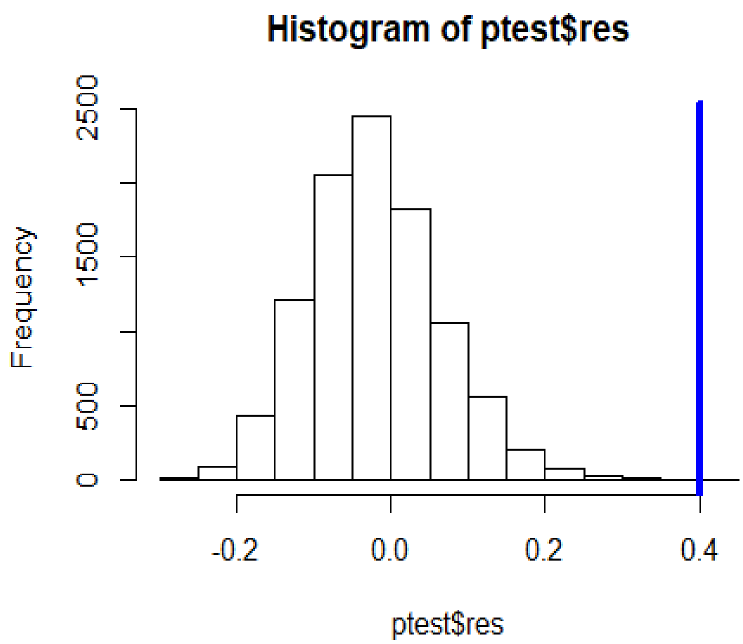


From the choropeth plot, It's clear that there is spatial correlation among the scores, as we can see clustering in the southwest and southeast, as well as correlation of higher scores in the midwest and northeast/New England areas. The QQ-plot indicates clustering in the middle and that the residuals are not i.i.d. as desired

b)

```
neighblist <- mat2listw(satprox)
#1000 permutation Moran's I test
ptest <- moran.mc(residslr, neighblist, nsim = 10000)

ptest
##
##  Monte-Carlo simulation of Moran I
##
## data:  residslr
## weights: neighblist
## number of simulations + 1: 10001
##
## statistic = 0.40151, observed rank = 10001, p-value = 9.999e-05
## alternative hypothesis: greater
Iactual <- as.numeric(ptest$res[10001])
Iactual
## [1] 0.4015076
hist(ptest$res)
abline(v = Iactual, col = "blue", lwd = 4)
```



Moran's I results show that there is strong evidence to reject the null hypothesis, indicating spatial correlation. The p-value is very small, and from the histogram we can see that the actual computed value is far in the tail of the distribution of the bootstrapped values.

c)

```
#spatial auto model, including intercept
spfit <- spautolm(formula = vscore ~ pc, listw = neighblist, family = "CAR", data=sat)
#summary of results
summary(spfit)
##
## Call: spautolm(formula = vscore ~ pc, data = sat, listw = neighblist,
##     family = "CAR")
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -25.30579  -7.15782   0.70458   9.80801  19.75856
##
## Coefficients:
##               Estimate Std. Error z value  Pr(>|z|)
## (Intercept) 575.019385   5.884447  97.719 < 2.2e-16
## pc           -1.094063   0.097837 -11.182 < 2.2e-16
##
## Lambda: 0.13853 LR test value: 11.529 p-value: 0.00068514
## Numerical Hessian standard error of lambda: 0.013629
##
## Log likelihood: -191.9149
## ML residual variance (sigma squared): 159.5, (sigma: 12.629)
## Number of observations: 48
## Number of parameters estimated: 4
## AIC: 391.83
```

d)

```
#now trying with SAR
sarfit <- spautolm(formula = vscore ~ pc, listw = neighblist, family = "SAR", data=sat)
#summary of results, very little difference from CAR model
summary(sarfit)
##
## Call: spautolm(formula = vscore ~ pc, data = sat, listw = neighblist,
##     family = "SAR")
##
## Residuals:
```

```
##       Min        1Q    Median        3Q       Max
## -25.53593  -7.52620  0.62487  10.35420  20.45613
##
## Coefficients:
##                Estimate Std. Error z value  Pr(>|z|)
## (Intercept) 574.663911   6.017124  95.505 < 2.2e-16
## pc           -1.097028   0.098767 -11.107 < 2.2e-16
##
## Lambda: 0.12324 LR test value: 15.804 p-value: 7.0247e-05
## Numerical Hessian standard error of lambda: 0.019577
##
## Log likelihood: -189.7774
## ML residual variance (sigma squared): 142.03, (sigma: 11.917)
## Number of observations: 48
## Number of parameters estimated: 4
## AIC: 387.55
```

The SAR model is very similar to the CAR model. The parameter estimates are very close to the same and the inferences on coefficients and parameters are not different. The AIC for the SAR model is smaller, so perhaps it is better, but the model does not change much.

```
#A new residual plot looks like we have accounted for much more of the spatial correlation
#though it may not be perfect yet
car.resids <- spfit$fit$residuals
plot(0, 0, xlim=c(-125, -68), ylim=c(25,50), type="n",
     xlab="", ylab="", main="residuals after accounting for spatial correlation")
plot.poly(US, car.resids[where.is.state], seq(from = -26, to = 20,by = 2),
          legend.x=c(-71, -70), legend.y=c(25, 30), add=TRUE)
```