Ward_Abigail_HW3

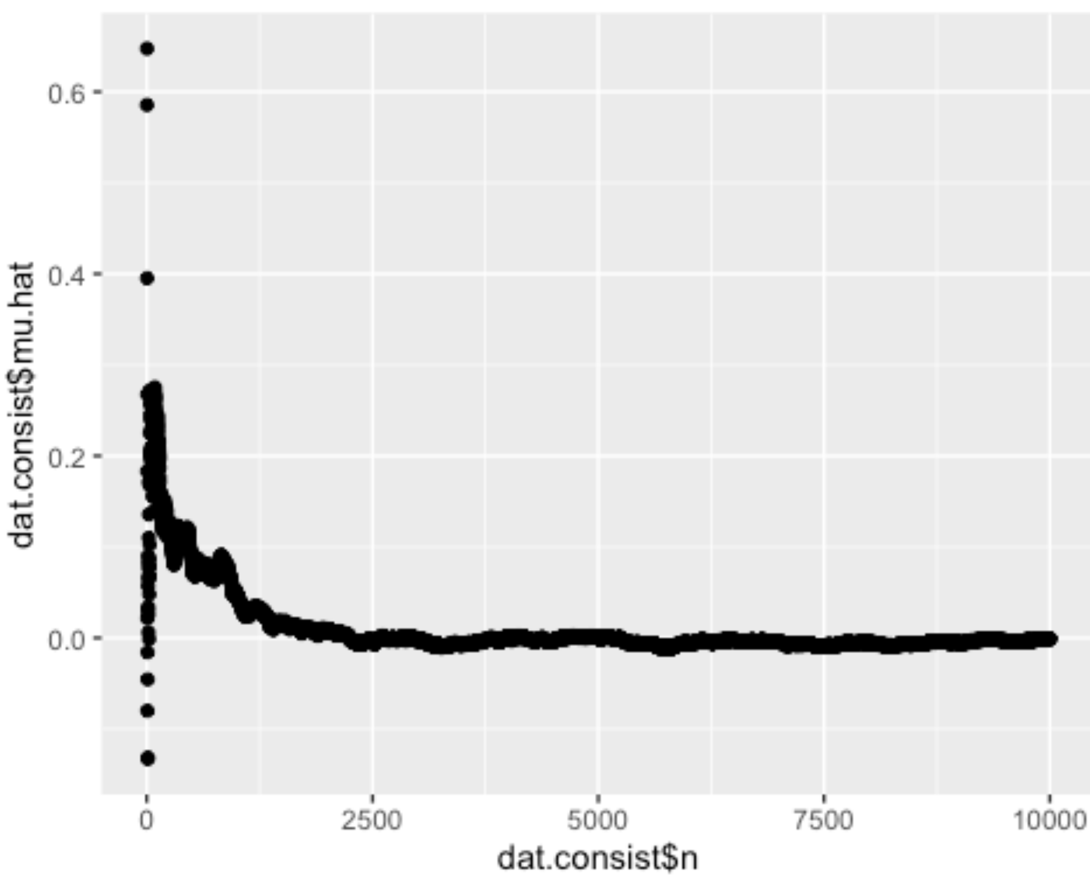
Abigail Ward

7/06/2021

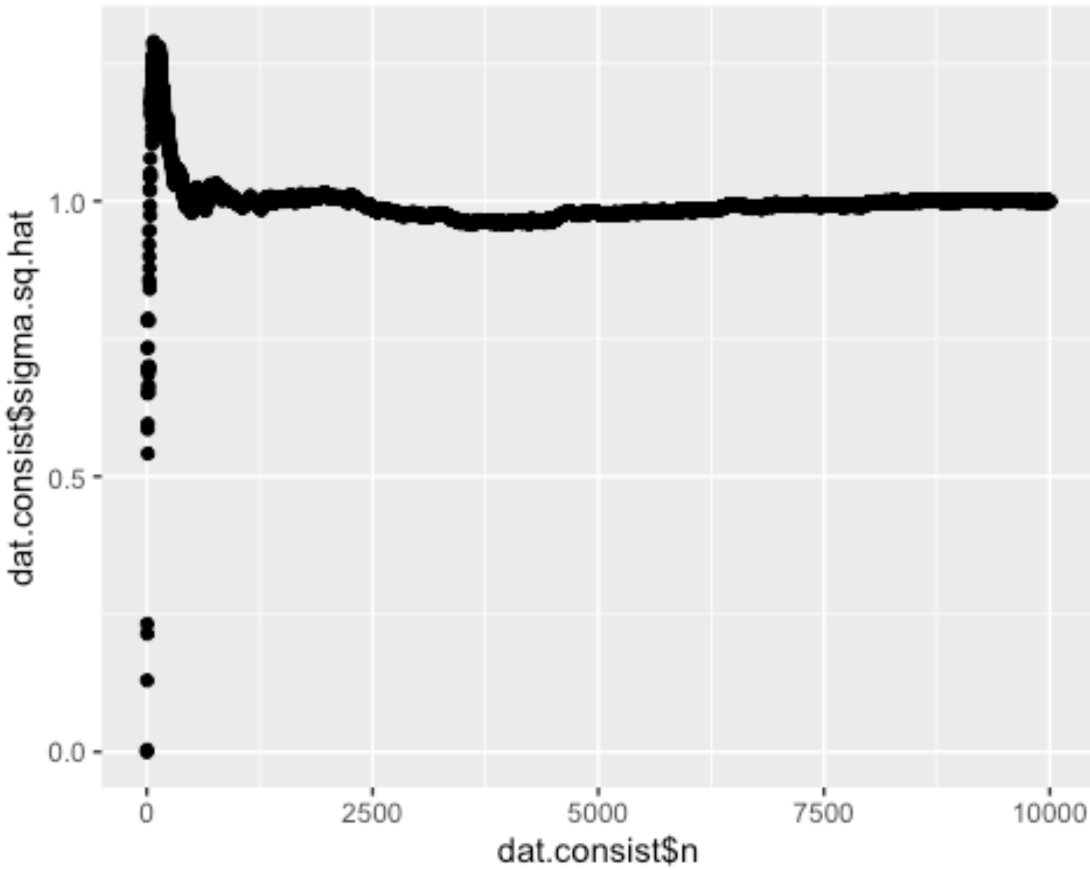
Question 1.a

In the plots below, you can see that the maximum likelihood estimates of μ and σ^2 are consistent and approach $\mu = 0$ and $\sigma^2 = 1$ as n increases toward infinity.

```
## Warning: Use of `dat.consist$n` is discouraged. Use `n` instead.

## Warning: Use of `dat.consist$mu.hat` is discouraged. Use `mu.hat` instead.
```




```
## Warning: Use of `dat.consist$n` is discouraged. Use `n` instead.

## Warning: Use of `dat.consist$sigma.sq.hat` is discouraged. Use
`sigma.sq.hat`
## instead.
```



Question 1.b

The maximum likelihood estimate of μ does seem to be unbiased since the average of 10000 small samples is still very close to 0. The maximum likelihood estimate of σ seems to be biased since the expected value was 1, but the calculated value was 0.805. This bias was corrected by dividing the sum of the squared differences by 4 instead of 5.

```
## [1] 0.008816334

## [1] 0.8049875

## [1] 1.006234
```

Question 2

$$f(\lambda) = \prod_{i=1}^{n} \lambda \exp(-\lambda x_i)$$

$$\ln f(\lambda) = \ln(\prod_{i=1}^{n} \lambda \exp(-\lambda x_i))$$

$$\ln f(\lambda) = \sum_{i=1}^{n} \ln(\lambda \exp(-\lambda x_i))$$

$$\ln f(\lambda) = \sum_{i=1}^{n} [ln\lambda + \ln(\exp(-\lambda x_i))]$$

$$\ln f(\lambda) = \sum_{i=1}^{n} \ln\lambda - \sum_{i=1}^{n} \lambda x_i$$

$$g(\lambda) = \ln f(\lambda) = n\lambda - \lambda \sum_{i=1}^{n} x_i$$

$$\frac{dg}{d\lambda} = n\frac{1}{\lambda} - \sum_{i=1}^{n} x_i = 0$$

$$\frac{1}{\lambda} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$$

$$\lambda = \frac{1}{\bar{x}}$$

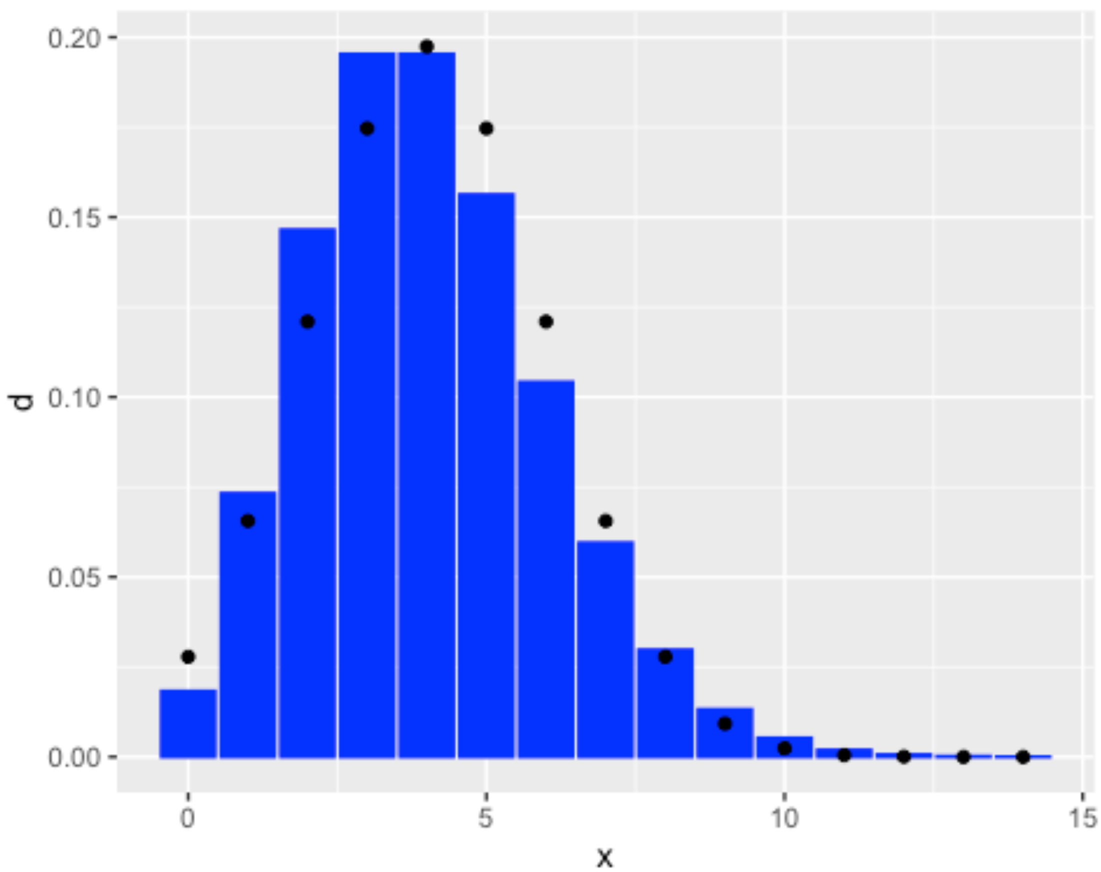
Question 3.a

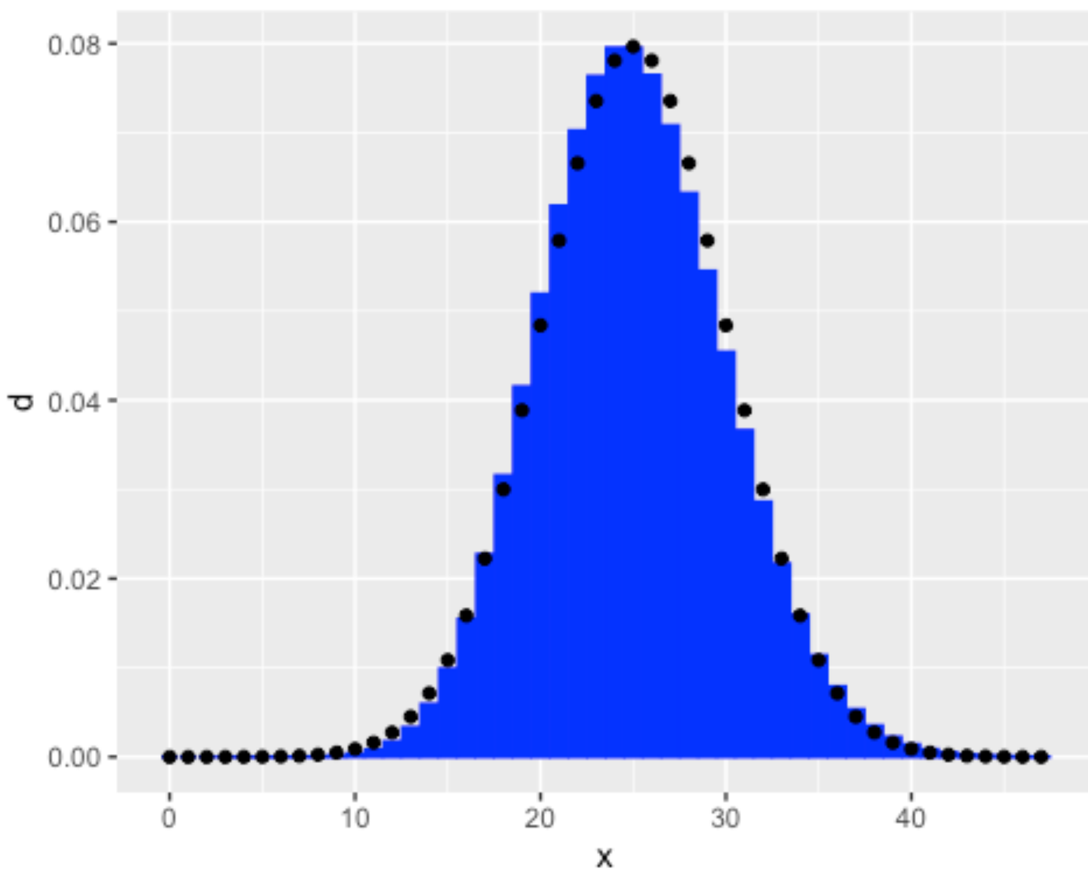
The maximum likelihood estimate of both μ and σ^2 are equal to the given λ

```
## [1]  3.99768 25.00485 99.98159

## [1]  3.997375 25.111318 99.068502
```

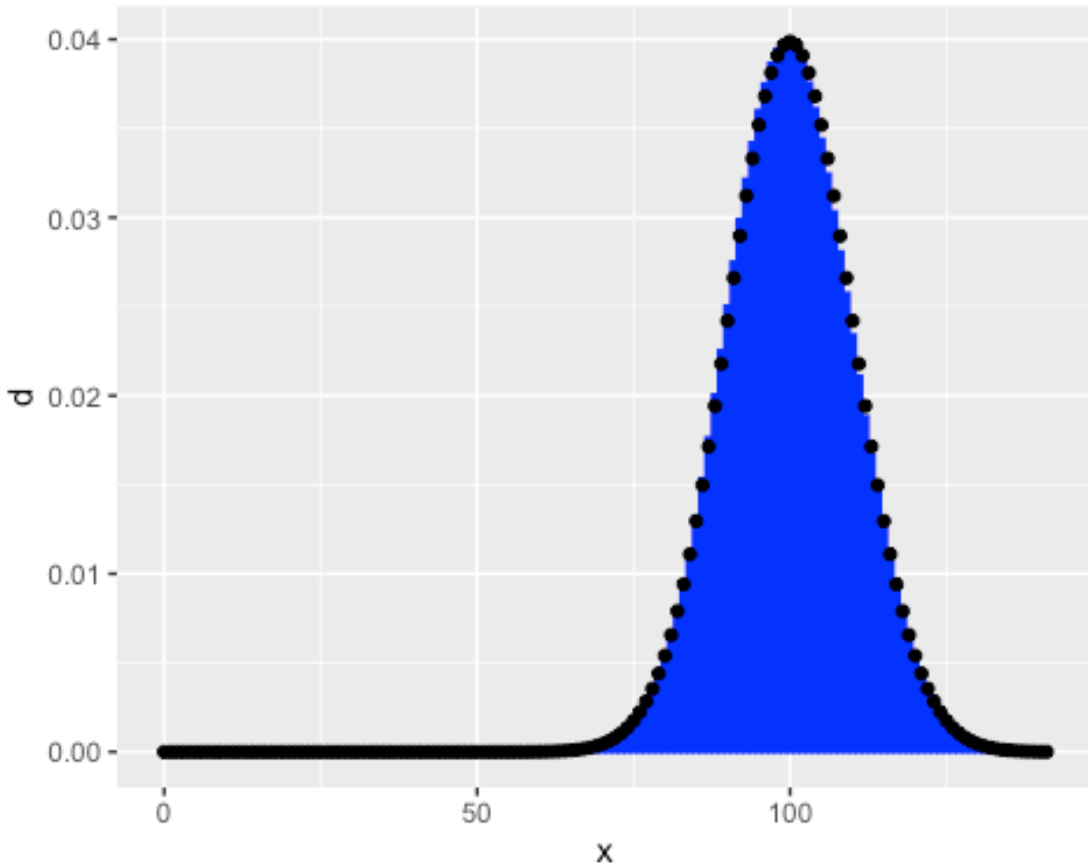
Question 3.b

The proposed Normal approximation to the $Poisson(\lambda)$ improves as λ increases as shown in the histograms better matching the normal distribution dots as well as the smaller sum of the absolute differences of the two probabilities.


0.20
0.15
0.10
0.05
0.00
d
0
5
10
15
x



0.08
0.06
0.04
0.02
0.00
d
0
10
20
30
40
x

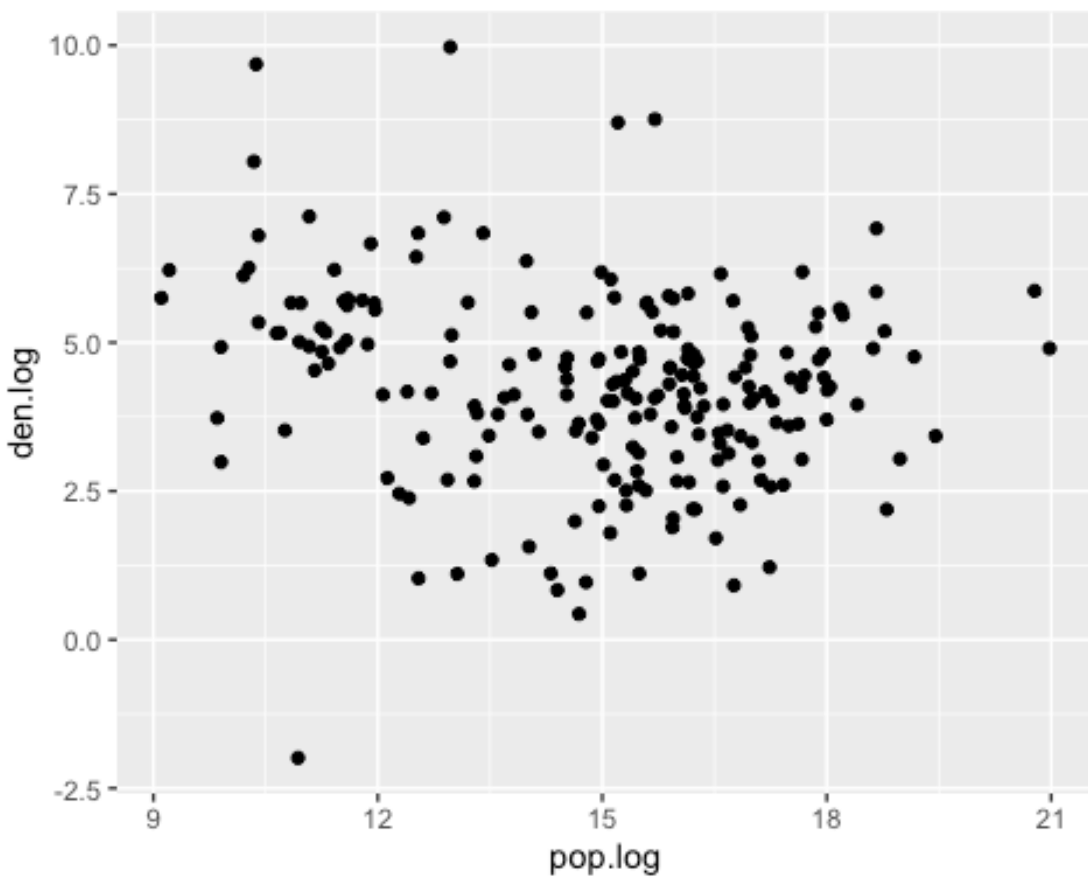



```
## [1] 0.1177445
## [1] 0.05049299
## [1] 0.02515794
```

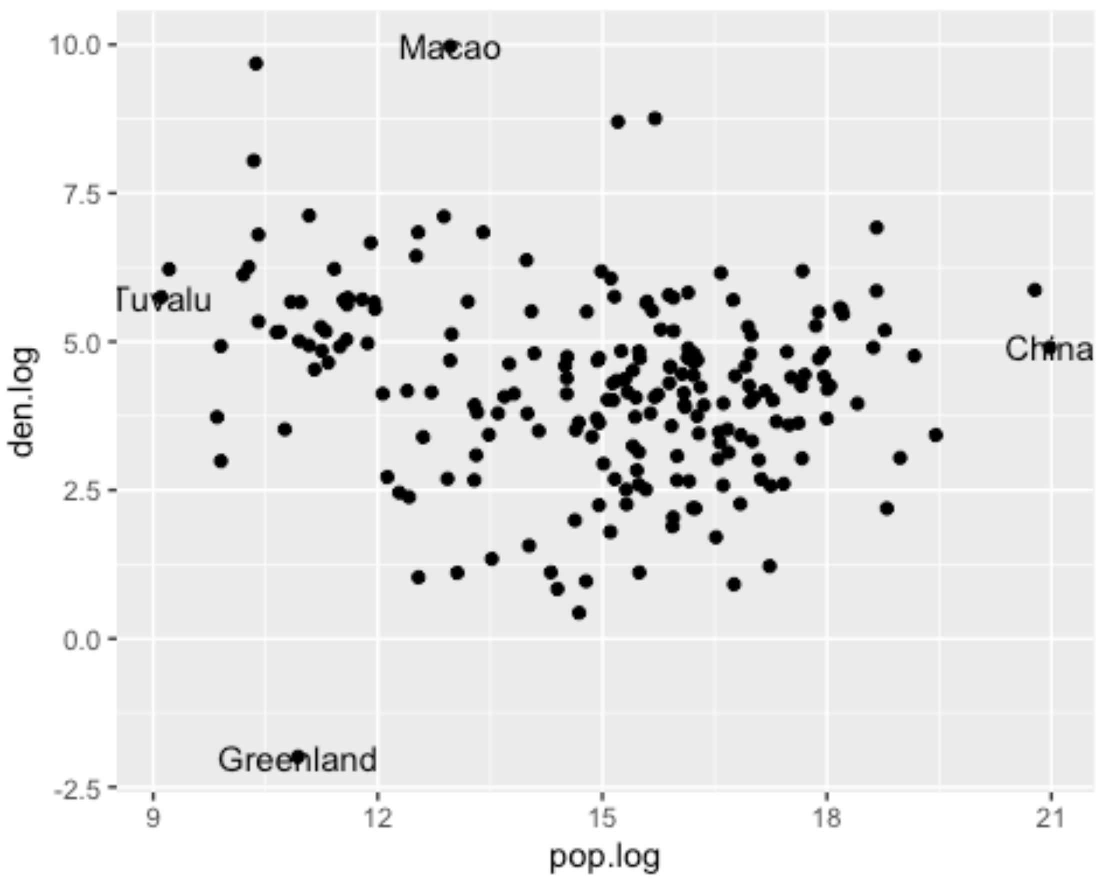
Question 4

```
dat.pop<-read.csv("population.csv",stringsAsFactors = FALSE)
dat.den<-
  read.csv("population-density.csv",stringsAsFactors = FALSE)
names(dat.den)[4]<-"density"

#4a answer
dat.pop<-filter(dat.pop,Year==2000,!Code %in% c("","OWID_WRL"))
dat.den<-filter(dat.den,Year==2000,!Code %in% c("","OWID_WRL"))


dat.both<-inner_join(dat.den,dat.pop,by="Code")
mean(dat.both$Entity.x==dat.both$Entity.y)

## [1] 1

#this is equal to 1
```

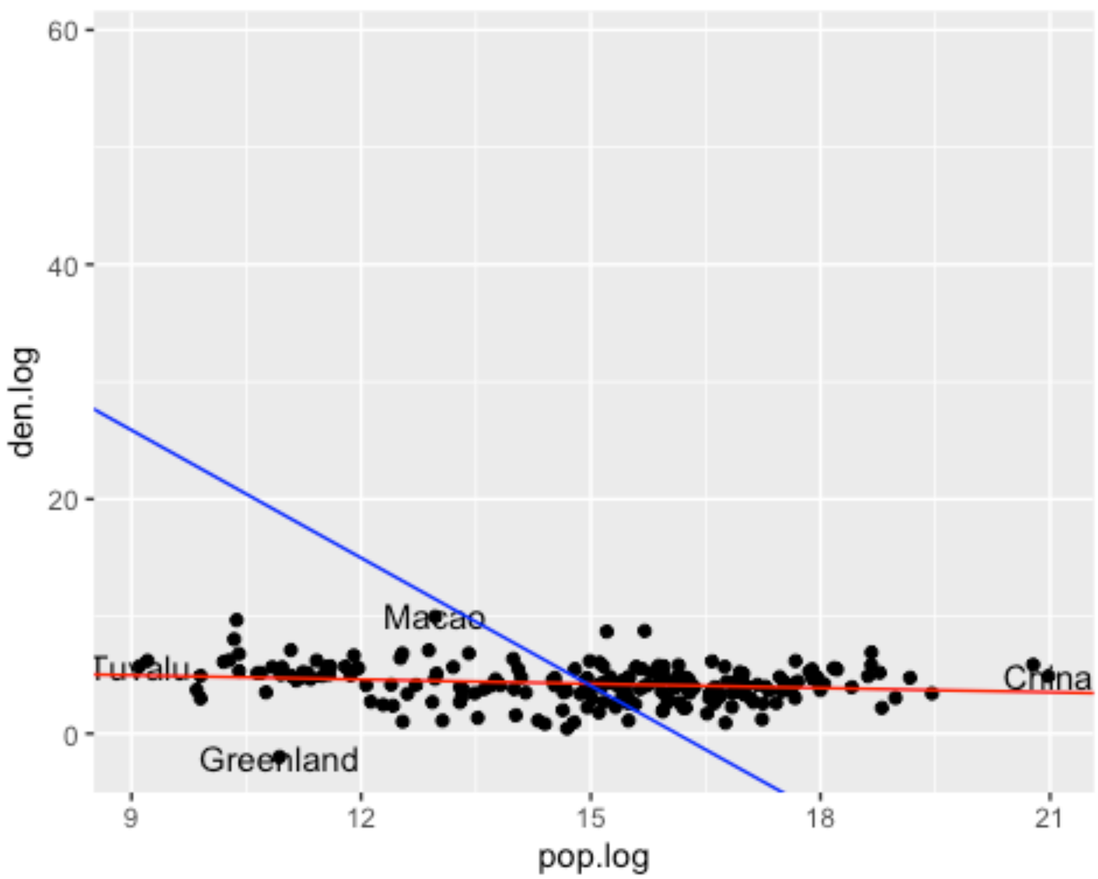
```
inds<-c(
  which(dat.both$density %in%
              c(max(dat.both$density),min(dat.both$density))),
  which(dat.both$Population %in%
            c(max(dat.both$Population),min(dat.both$Population)))
)

dat.both<-transmute(dat.both,den.log=log(density),
               pop.log=log(Population),entity=Entity.x)


dat.text<-dat.both[inds,]

g<-ggplot(dat.both,aes(x=pop.log,y=den.log))+geom_point()
g
```




```
g<-g+
  geom_text(data=dat.text,aes(x=pop.log,y=den.log,label=entity))
g
```




```
lin.model<-lm(den.log~pop.log, data=dat.both)

coef<-lm(den.log~pop.log,data=dat.both)$coefficients
coef

## (Intercept)     pop.log
##   6.0776440  -0.1214985

coefRev<-lm(pop.log~den.log,data=dat.both)$coefficients
slopeRev<-1/coefRev[2]
interceptRev<- (-coefRev[1])/coefRev[2]

g<-g+
  geom_abline(slope=coef[2],intercept=coef[1],color="red") +
  geom_abline(slope=slopeRev,intercept=interceptRev,
              color="blue")
g
```


60
40
20
0
den.log
Macao
Tuvalu
China
Greenland
9
12
15
18
21
pop.log