Geography of Birding in the United States

Jacob C. Cooper

August 28, 2019

Introduction

eBird data was provided by Ian Davies in February 2019. All other data were gleaned from Wikipedia in February 2019. GDP was taken from the Wikipedia page on GDP. These codes also include a "distance to line" function that was adopted from a StackOverflow discussion.

First, we will load required packages.

```
library(ggplot2)
```

Note that the data were aligned manually in a spreadsheet to ensure all county names were consistent.

Analyses

Load the data

First, we need to load the data table.

```
x=read.csv(paste0(filepath,"GDP-eBird_data.csv"))

head(x)
```

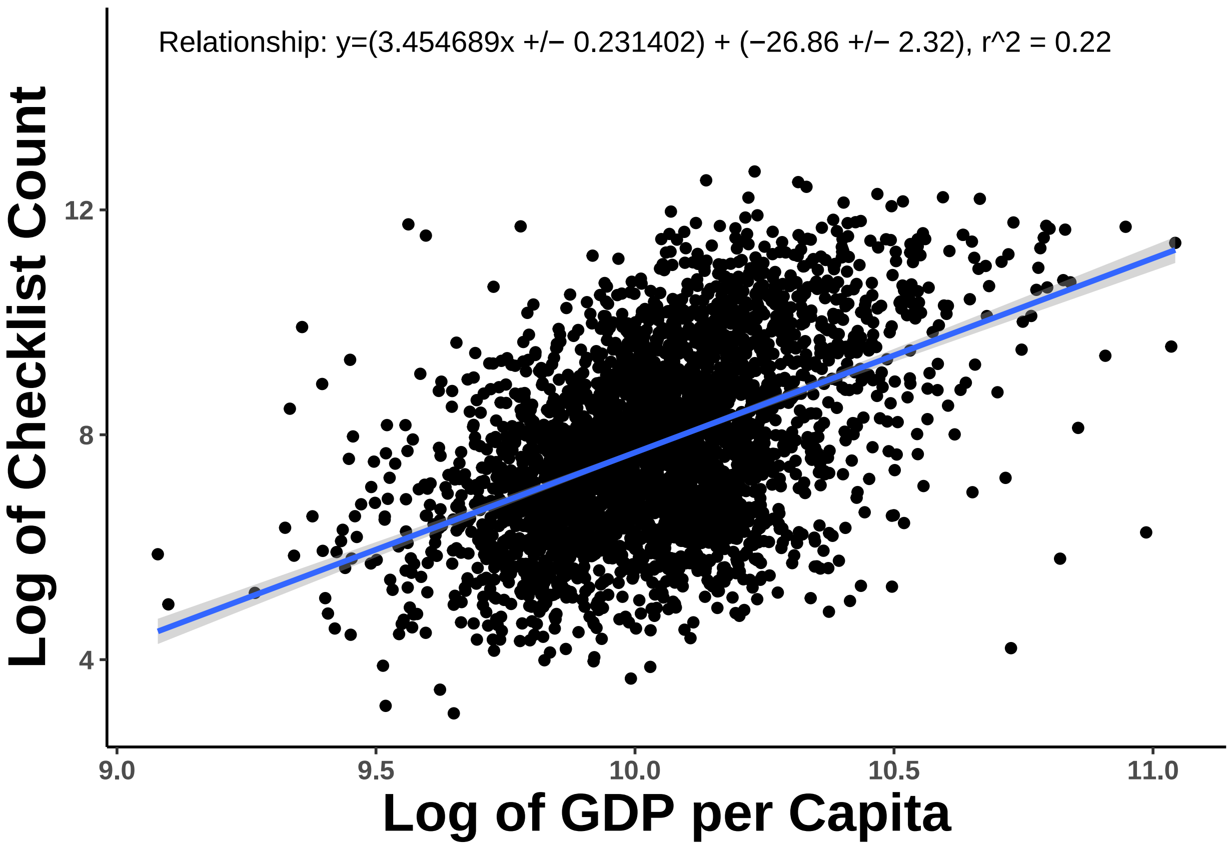
```
##   eBird_count eBird_code   State  County PerCapitaIncome Population
## 1        1183  US-AL-001 Alabama Autauga           24571      54907
## 2       26908  US-AL-003 Alabama Baldwin           26766     187114
## 3        2862  US-AL-005 Alabama Barbour           16829      27321
## 4         639  US-AL-007 Alabama    Bibb           17427      22754
## 5         899  US-AL-009 Alabama  Blount           20730      57623
## 6        2907  US-AL-011 Alabama Bullock           18628      10746
```

GDP Per Capita

First, we can compare the per capita income of each individual area to the number of checklists accumulated in that area.

```
# add column for natural log of checklist count
x$logcheck=log(x$eBird_count)

# add column for natural log of GDP
x$logGDP=log(x$PerCapitaIncome)

# get regression equation
eq.x.1=lm(x$logcheck~x$logGDP)
a.1.1=round(coef(eq.x.1)[1],2)#intercept
b.1.1=round(coef(eq.x.1)[2],6)#slope
r.x.1=round(summary(eq.x.1)$r.squared,2)
```

```
a.err.1=round(summary(eq.x.1)$coefficients[3],2)
b.err.1=round(summary(eq.x.1)$coefficients[4],6)

# create plot
a=ggplot(aes(x=logGDP,y=logcheck),data=x)
b=geom_point()
c=geom_smooth(method=lm)
d=theme_classic()
d.1=theme(axis.title = element_text(face="bold",size=20),
          axis.text = element_text(size=10,face="bold"),
          legend.title = element_blank(),
          legend.text = element_text(size=14,face="bold"))

e=labs(x='Log of GDP per Capita',y='Log of Checklist Count')
p=annotate("text",label=paste0('Relationship: y=(',b.1.1,'x +/- ',2*b.err.1,') + (',a.1.1," +/- ",2*a.e
#p.1=annotate("text",label=paste0('P. villosus: y=(',b.1,'x +/- ',2*b.err,') + (',a.1," +/- ",2*a.err,'
Fig3=a+b+c+d+e+d.1+p

plot(Fig3)
```

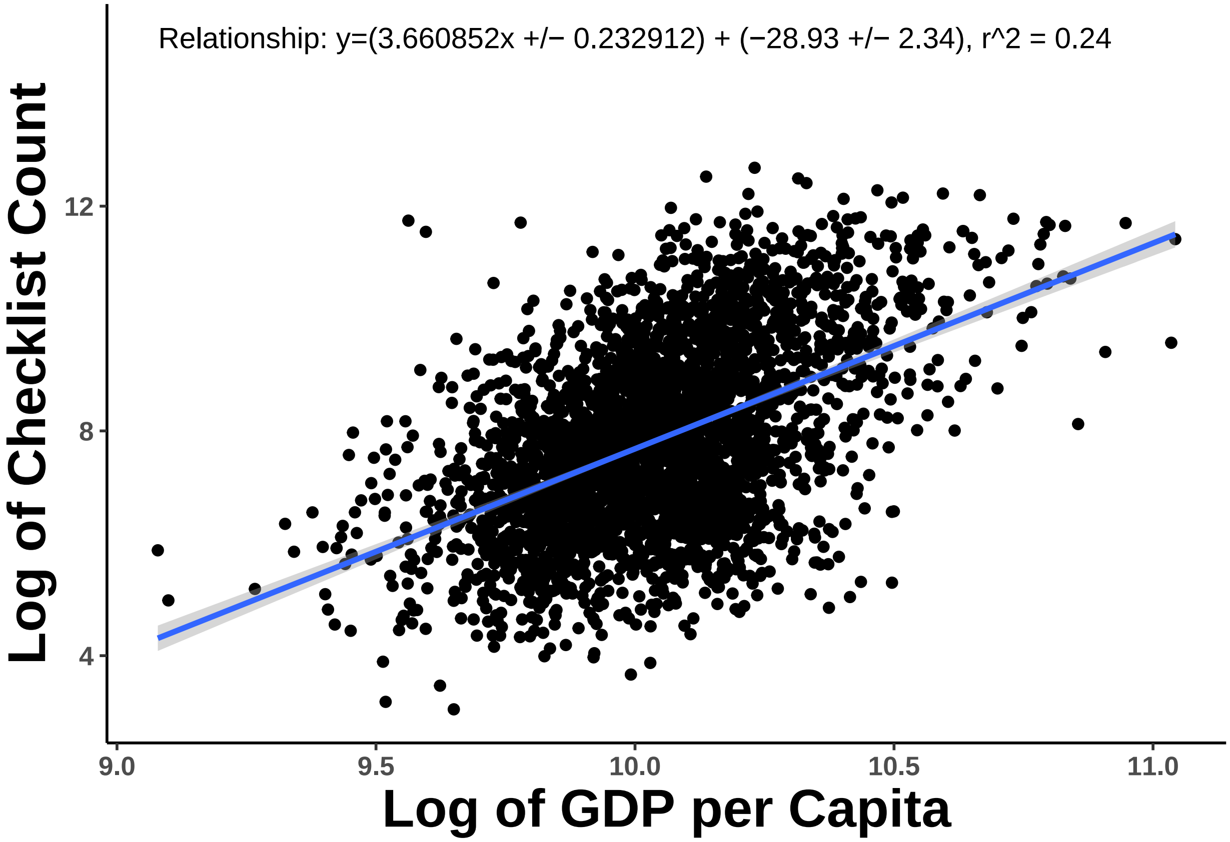


There is a fairly amorphous cloud of points, and that the relationship within the data is not extremely strong ($R^2 = 0.22$).

We can try to clip out extreme variation to see if this improves the fit.

```
subset1=which(x$logcheck<8&x$logGDP>10.5)
subset2=which(x$logcheck>8&x$logGDP<9.5)
```

```
subset=x[-c(subset1,subset2),]

eq.x.1=lm(subset$logcheck~subset$logGDP)
a.1.1=round(coef(eq.x.1)[1],2)#intercept
b.1.1=round(coef(eq.x.1)[2],6)#slope
r.x.1=round(summary(eq.x.1)$r.squared,2)
a.err.1=round(summary(eq.x.1)$coefficients[3],2)
b.err.1=round(summary(eq.x.1)$coefficients[4],6)

a=ggplot(aes(x=logGDP,y=logcheck),data=subset)
b=geom_point()
c=geom_smooth(method=lm)
d=theme_classic()
d.1=theme(axis.title = element_text(face="bold",size=20),
          axis.text = element_text(size=10,face="bold"),
          legend.title = element_blank(),
          legend.text = element_text(size=14,face="bold"))

e=labs(x='Log of GDP per Capita',y='Log of Checklist Count')
p=annotate("text",label=paste0('Relationship: y=(',b.1.1,'x +/- ',2*b.err.1,') + (',a.1.1," +/- ",2*a.e
#p.1=annotate("text",label=paste0('P. villosus: y=(',b.1,'x +/- ',2*b.err,') + (',a.1," +/- ",2*a.err,'
Fig3=a+b+c+d+e+d.1+p

plot(Fig3)
```

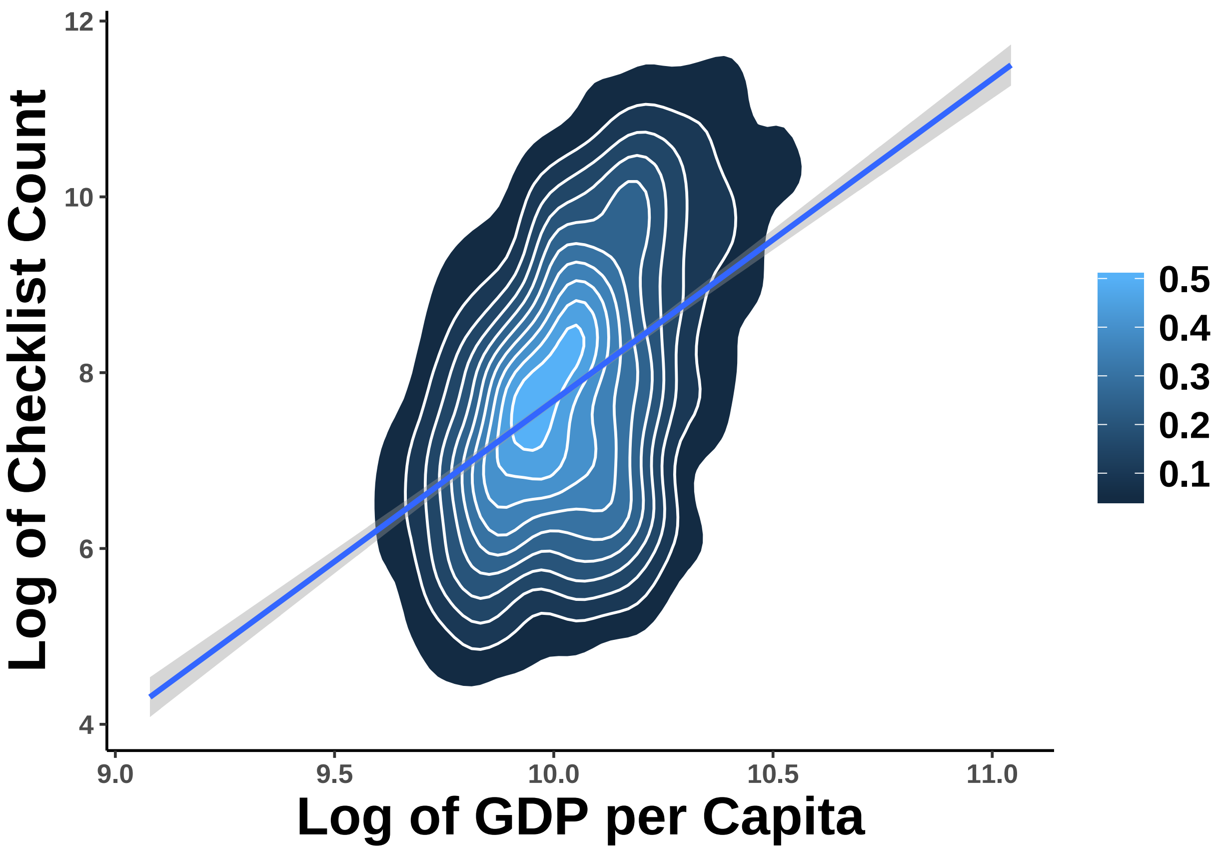


It fits better, but not by a lot. There is still a lot of uncertainty in the cloud center. Let's look at the point density to determine what may be causing this.

```r
a=ggplot(aes(x=logGDP,y=logcheck),data=subset)
b=stat_density_2d(aes(fill=..level..),geom="polygon",colour="white")
c=geom_smooth(method=lm)
d=theme_classic()
d.1=theme(axis.title = element_text(face="bold",size=20),
          axis.text = element_text(size=10,face="bold"),
          legend.title = element_blank(),
          legend.text = element_text(size=14,face="bold"))

e=labs(x='Log of GDP per Capita',y='Log of Checklist Count')
#p=annotate("text",label=paste0('Relationship: y=(',b.1.1,'x +/- ',2*b.err.1,') + (',a.1.1," +/- ",2*a.
#p.1=annotate("text",label=paste0('P. villosus: y=(',b.1,'x +/- ',2*b.err,') + (',a.1," +/- ",2*a.err,'
Fig3=a+b+c+d+e+d.1

paste0('Relationship: y=(',b.1.1,'x +/- ',2*b.err.1,') + (',
       a.1.1," +/- ",2*a.err.1,')',', r^2 = ',r.x.1)
```

```
## [1] "Relationship: y=(3.660852x +/- 0.232912) + (-28.93 +/- 2.34), r^2 = 0.24"
```

```r
plot(Fig3)
```



We have a fit of $R^2 = 0.24$, which is better but not by a lot.

As is visible in the density plot, the tails of the distribution are dragging it down and there is a secondary area of increased density on the lower part of the cluster. A closer look at this sub-group reveals the following:

```r
#Restrict GDP
sub2=subset[subset$logGDP>10&subset$logGDP<10.25,]
```

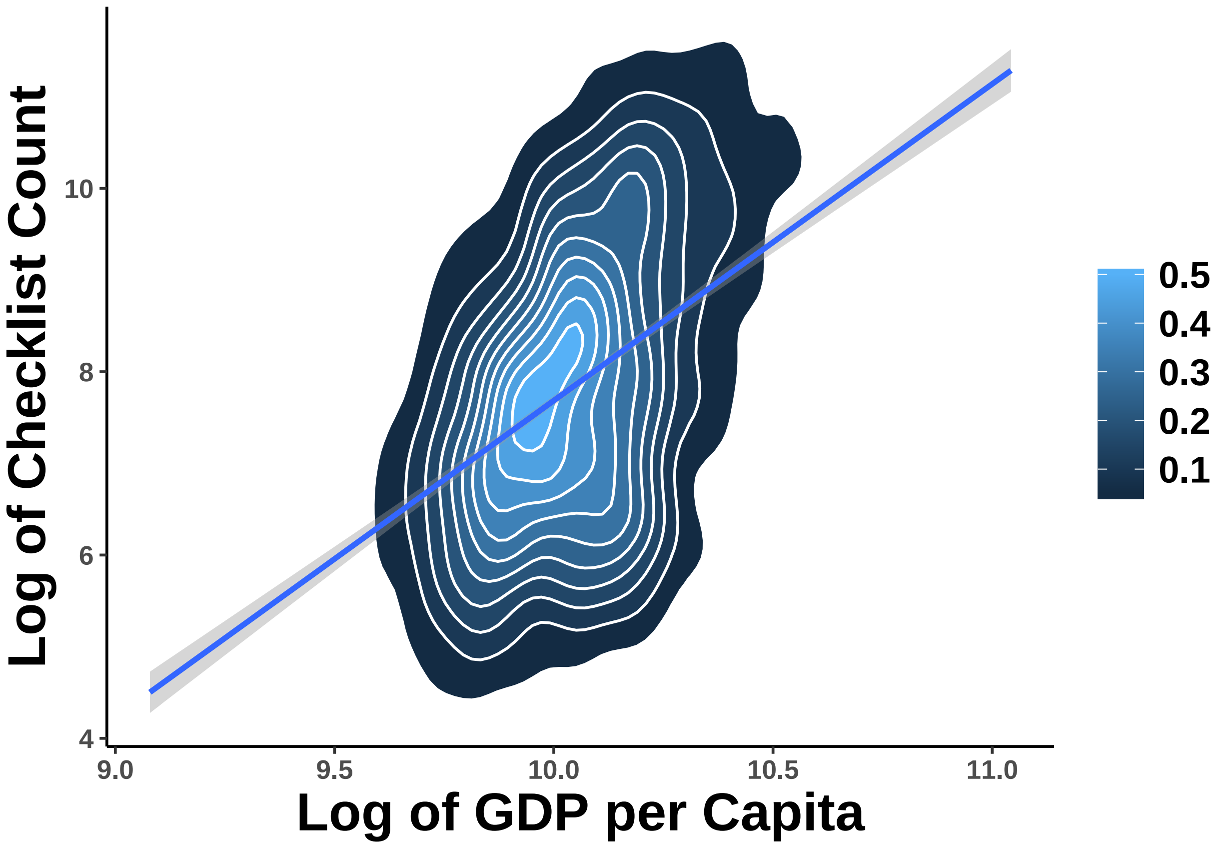
```r
#Restrict lists
sub2=sub2[sub2$logcheck<7&sub2$logcheck>6,]

summary(sub2)
```

```
##    eBird_count        eBird_code           State           County   
##  Min.   : 405.0   US-AL-031:  1   Kansas      : 34   Grant   :  4  
##  1st Qu.: 531.8   US-AR-105:  1   Nebraska    : 30   Hamilton:  4  
##  Median : 682.0   US-CO-033:  1   Iowa        : 28   Carroll :  3  
##  Mean   : 702.5   US-CO-111:  1   Texas       : 20   Clark   :  3  
##  3rd Qu.: 849.8   US-GA-097:  1   South Dakota: 18   Clay    :  3  
##  Max.   :1095.0   US-GA-103:  1   Illinois    : 15   Hancock :  3  
##                   (Other)  :250   (Other)     :111   (Other) :236  
##  PerCapitaIncome   Population        logcheck         logGDP     
##  Min.   :22035   Min.   :   659   Min.   :6.004   Min.   :10.00  
##  1st Qu.:23357   1st Qu.:  5225   1st Qu.:6.276   1st Qu.:10.06  
##  Median :24642   Median :  9444   Median :6.525   Median :10.11  
##  Mean   :24808   Mean   : 16539   Mean   :6.515   Mean   :10.12  
##  3rd Qu.:26109   3rd Qu.: 20242   3rd Qu.:6.745   3rd Qu.:10.17  
##  Max.   :28241   Max.   :143845   Max.   :6.999   Max.   :10.25  
## 
```

This secondary cluster is made up of states with large agricultural areas - Nebraska, Iowa, Texas, South Dakota. Specifically, these are agricultural regions with similar incomes and similar low-level consistent checklist output. A look at each area by region will provide more insight.

Create official GDP plot

```r
# get regression equation
eq.x.1=lm(x$logcheck~x$logGDP)
a.1.1=round(coef(eq.x.1)[1],2)#intercept
b.1.1=round(coef(eq.x.1)[2],6)#slope
r.x.1=round(summary(eq.x.1)$r.squared,2)
a.err.1=round(summary(eq.x.1)$coefficients[3],2)
b.err.1=round(summary(eq.x.1)$coefficients[4],6)

# create plot
a=ggplot(aes(x=logGDP,y=logcheck),data=x)
b=stat_density_2d(aes(fill=..level..),geom="polygon",colour="white")
c=geom_smooth(method=lm)
d=theme_classic()
d.1=theme(axis.title = element_text(face="bold",size=20),
          axis.text = element_text(size=10,face="bold"),
          legend.title = element_blank(),
          legend.text = element_text(size=14,face="bold"))

e=labs(x='Log of GDP per Capita',y='Log of Checklist Count')
Fig3=a+b+c+d+e+d.1

plot(Fig3)
```




```
paste0('Relationship: y=(',b.1.1,'x +/- ',2*b.err.1,') + (',
       a.1.1," +/- ",2*a.err.1,')',', r^2 = ',r.x.1)
```

```
## [1] "Relationship: y=(3.454689x +/- 0.231402) + (-26.86 +/- 2.32), r^2 = 0.22"
```

```
ggsave(plot=Fig3,filename=paste0(filepath,"GDP-vs-Checklist_density.png"),dpi=400)
```

```
## Saving 6.5 x 4.5 in image
```

There is a fairly amorphous cloud of points, and that the relationship within the data is not extremely strong ($R^2 = 0.22$).

Regional Comparisons of GDP per Capita

The GDP exploration revealed clear patterns in terms of region. I will now subdivide the states by Census REgions and Divisions as defined by the USA government; note that at the more local scale I exclude Alaska and Hawaii given their geographic uniqueness.

```
x$Region="undefined"
x$Subregion="undefined"

#define subregions

NewEngland=which(as.character(x$State)=="Connecticut"|
                   as.character(x$State)=="Massachusetts"|
                   as.character(x$State)=="Maine"|
                   as.character(x$State)=="New Hampshire"|
                   as.character(x$State)=="Rhode Island"|
```

```
                   as.character(x$State)=="Vermont")

MidAtlantic=which(as.character(x$State)=="New York"|
                    as.character(x$State)=="New Jersey"|
                    as.character(x$State)=="Pennsylvania")

ENCentral=which(as.character(x$State)=="Ohio"|
                  as.character(x$State)=="Michigan"|
                  as.character(x$State)=="Indiana"|
                  as.character(x$State)=="Illinois"|
                  as.character(x$State)=="Wisconsin")

WNCentral=which(as.character(x$State)=="Minnesota"|
                  as.character(x$State)=="Iowa"|
                  as.character(x$State)=="Missouri"|
                  as.character(x$State)=="Kansas"|
                  as.character(x$State)=="Nebraska"|
                  as.character(x$State)=="South Dakota"|
                  as.character(x$State)=="North Dakota")

SAtlantic=which(as.character(x$State)=="Delaware"|
                  as.character(x$State)=="Maryland"|
                  as.character(x$State)=="District of Columbia"|
                  as.character(x$State)=="West Virginia"|
                  as.character(x$State)=="Virginia"|
                  as.character(x$State)=="South Carolina"|
                  as.character(x$State)=="North Carolina"|
                  as.character(x$State)=="Georgia"|
                  as.character(x$State)=="Florida")

ESCentral=which(as.character(x$State)=="Kentucky"|
                  as.character(x$State)=="Tennessee"|
                  as.character(x$State)=="Mississippi"|
                  as.character(x$State)=="Alabama")

WSCentral=which(as.character(x$State)=="Arkansas"|
                  as.character(x$State)=="Louisiana"|
                  as.character(x$State)=="Oklahoma"|
                  as.character(x$State)=="Texas")

Mountain=which(as.character(x$State)=="Montana"|
                 as.character(x$State)=="Idaho"|
                 as.character(x$State)=="Wyoming"|
                 as.character(x$State)=="Colorado"|
                 as.character(x$State)=="Utah"|
                 as.character(x$State)=="Nevada"|
                 as.character(x$State)=="Arizona"|
                 as.character(x$State)=="New Mexico")

Pacific=which(as.character(x$State)=="Washington"|
                as.character(x$State)=="Oregon"|
                as.character(x$State)=="California")
```

```
x$Subregion[NewEngland]="New England"
x$Subregion[MidAtlantic]="Middle Atlantic"
x$Subregion[SAtlantic]="South Atlantic"
x$Subregion[ESCentral]="East South Central"
x$Subregion[WSCentral]="West South Central"
x$Subregion[ENCentral]="East North Central"
x$Subregion[WNCentral]="West North Central"
x$Subregion[Mountain]="Mountain"
x$Subregion[Pacific]="Pacific"

x$Region[c(NewEngland,MidAtlantic)]="Northeast"
x$Region[c(SAtlantic,ESCentral,WSCentral)]="South"
x$Region[c(ENCentral,WNCentral)]="Midwest"
x$Region[c(Mountain,Pacific)]="West"

x$Region=as.factor(x$Region)
x$Subregion=as.factor(x$Subregion)
```

First, by major region:

```
regions=unique(x$Region) #remove undefined

for(i in 1:length(regions)){
  subset.x=x[x$Region==regions[i],]
  subset.x=na.omit(subset.x)

  eq.x.1=lm(subset.x$logcheck~subset.x$logGDP)
  a.1.1=round(coef(eq.x.1)[1],2)#intercept
  b.1.1=round(coef(eq.x.1)[2],6)#slope
  r.x.1=round(summary(eq.x.1)$r.squared,2)
  a.err.1=round(summary(eq.x.1)$coefficients[3],2)
  b.err.1=round(summary(eq.x.1)$coefficients[4],6)

  xval=min(subset.x$logGDP)+(max(subset.x$logGDP)-min(subset.x$logGDP))/2
  yval=1+max(subset.x$logcheck)

  a=ggplot(aes(x=logGDP,y=logcheck),data=subset.x)
  b=geom_point()
  c=geom_smooth(method=lm)
  d=theme_classic()
  d.1=theme(axis.title = element_text(face="bold",size=20),
            axis.text = element_text(size=10,face="bold"),
            legend.title = element_blank(),
            legend.text = element_text(size=14,face="bold"))

  e=labs(x='Log of GDP per Capita',y='Log of Checklist Count')
  w2=ggtitle(regions[i])

  Fig3=a+b+c+d+e+d.1+w2

  ggsave(plot=Fig3,filename=paste0(filepath,"GDP-vs-Checklist_region_",regions[i],".png"),dpi=400)

  plot(Fig3)
  print(paste0('Relationship: ',regions[i],' y=(',round(b.1.1,3),'x +/- ',round(2*b.err.1,3),
```

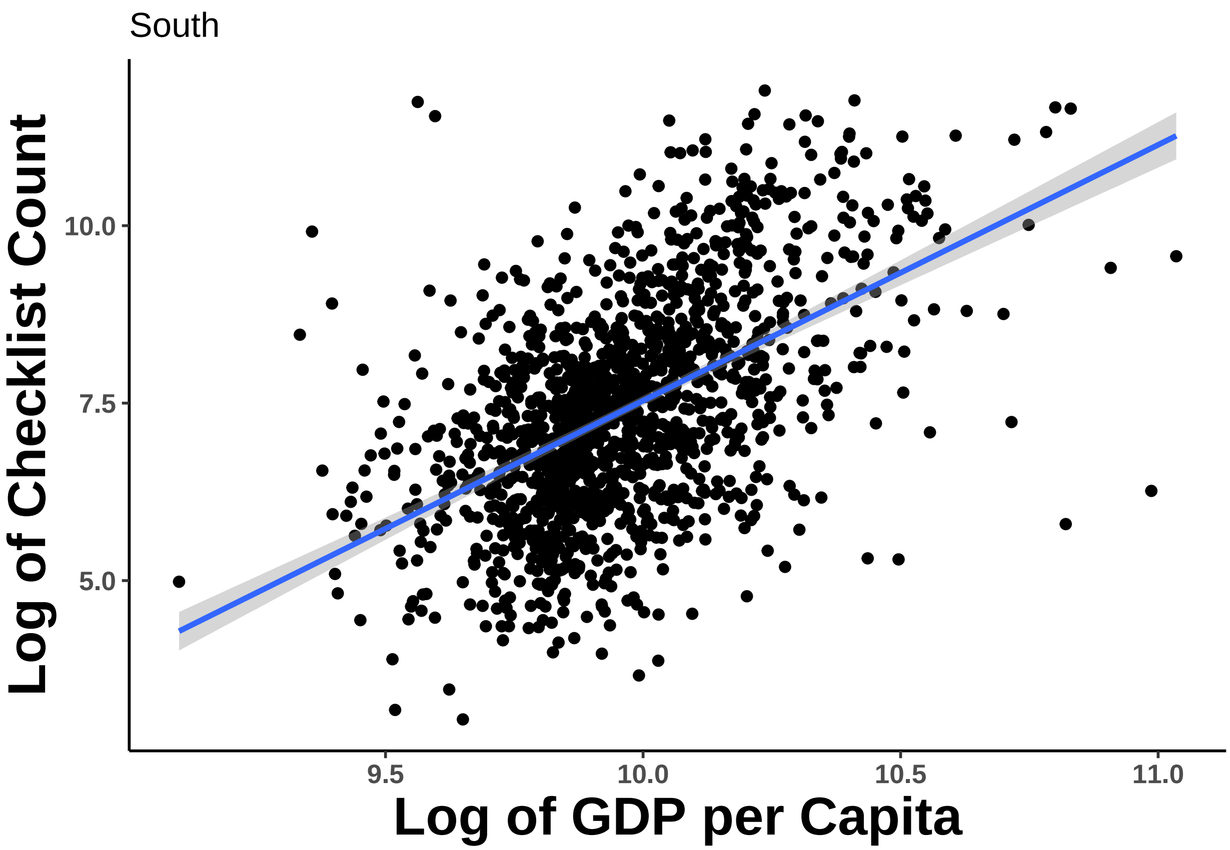
```
                                   ') + (',round(a.1.1,3)," +/- ",round(2*a.err.1,3),')',
                                   ', r^2 = ',round(r.x.1,2)))
  #What are the biggest residual outliers?

  dist2d=function(a1,a2,a3){
    v1=a2-a3
    v2=a1-a2
    m=cbind(v1,v2)
    d=det(m)/sqrt(sum(v1*v1))
    return(d)
  }

  a2=c(0,a.1.1)
  val.test=(b.1.1*20)+a.1.1
  a3=c(20,val.test)

  subset.x$RDistance=0

  for(j in 1:nrow(subset.x)){
    a1.x=subset.x$logGDP[j]
    a1.y=subset.x$logcheck[j]
    a1=c(a1.x,a1.y)
    subset.x$RDistance[j]=dist2d(a1,a2,a3)
  }

  mu=mean(subset.x$RDistance)
  sd.x=sd(subset.x$RDistance)

  hi=qnorm(p=0.975,mean=mu,sd=sd.x)
  lo=qnorm(p=0.025,mean=mu,sd=sd.x)

  lows=subset.x[subset.x$RDistance<lo,]
  his=subset.x[subset.x$RDistance>hi,]

  lows=lows[order(lows$RDistance,decreasing=F),]
  his=his[order(his$RDistance,decreasing=T),]

  #Note that somehow hi and low are switched

  print("Higher than Average Counties")
  print(lows[,c("County","State","PerCapitaIncome","eBird_count")])

  print("Lower than Average Counties")
  print(his[,c("County","State","PerCapitaIncome","eBird_count")])
}
```

```
## Saving 6.5 x 4.5 in image

## [1] "Relationship: South y=(3.605x +/- 0.309) + (-28.51 +/- 3.08), r^2 = 0.28"
## [1] "Higher than Average Counties"
##                County          State PerCapitaIncome eBird_count
## 2628          Hidalgo          Texas        14222.00      125839
## 2551          Cameron          Texas        14710.00      103051
## 2734            Starr          Texas        11584.00       20262
```

```
## 361        Miami-Dade        Florida        23174.00        96825
## 2473             Lake      Tennessee        12042.00         7354
## 2621           Harris          Texas        27899.00       147841
## 2765          Willacy          Texas        11313.00         4739
## 2643        Jefferson          Texas        23236.00        61919
## 353               Lee        Florida        27348.00       105912
## 319           Alachua        Florida        24857.00        74385
## 2698           Nueces          Texas        23671.00        61192
## 2355      Spartanburg South Carolina        21889.00        45336
## 414            Clarke        Georgia        19295.00        28420
## 2535            Bexar          Texas        24253.00        63563
## 323           Brevard        Florida        27009.00        92474
## 371              Polk        Florida        21285.00        35764
## 366            Orange        Florida        24877.00        62262
## 2760             Webb          Texas        14553.00         8816
## 2920          Radford       Virginia        16181.00        12759
## 416           Clayton        Georgia        17950.00        17675
## 2818         Accomack       Virginia        22703.00        38487
## 1194 BaltimoreCity+Co       Maryland        30202.44       104013
## 367           Osceola        Florida        19007.00        19590
## 370          Pinellas        Florida        29262.00        91724
## 317            Sussex       Delaware        26908.00        64511
## 2651           Kenedy          Texas        15157.00         7669
## 2747           Travis          Texas        33206.00       128626
## 1209         Somerset       Maryland        16748.00        10588
## 2916    Prince Edward       Virginia        17208.00        11642
## 2604        Galveston          Texas        30926.00        95870
## 315              Kent       Delaware        24851.00        42148
## 333             Duval        Florida        26143.00        49184
## 343           Hendry        Florida        16133.00         8252
## 2340           Jasper South Carolina        17350.00        10349
## 344          Hernando        Florida        21411.00        22046
## 49             Mobile        Alabama        22501.00        26272
## [1] "Lower than Average Counties"
##             County         State PerCapitaIncome eBird_count
## 2857  Falls Church      Virginia           59088         525
## 2537        Borden         Texas           50042         329
## 2717       Roberts         Texas           36172         200
## 3042          Wirt West Virginia           21852          39
## 2671        Loving         Texas           34068         203
## 1037       Hancock      Kentucky           22686          48
## 2894 Manassas Park      Virginia           26944         119
## 1417      Franklin   Mississippi           24234          93
## 2668      Lipscomb         Texas           29017         180
## 1463         Smith   Mississippi           20334          53
## 1068        Martin      Kentucky           15525          21
## 2906        Norton      Virginia           22699          92
## 1103       Trimble      Kentucky           22070          95
## 2638         Irion         Texas           28055         226
## 1437      Lawrence   Mississippi           20655          79
## 1468        Tippah   Mississippi           18493          54
## 2655          King         Texas           29836         304
## 2868     Goochland      Virginia           45039        1385
## 1471         Union   Mississippi           19273          66
```

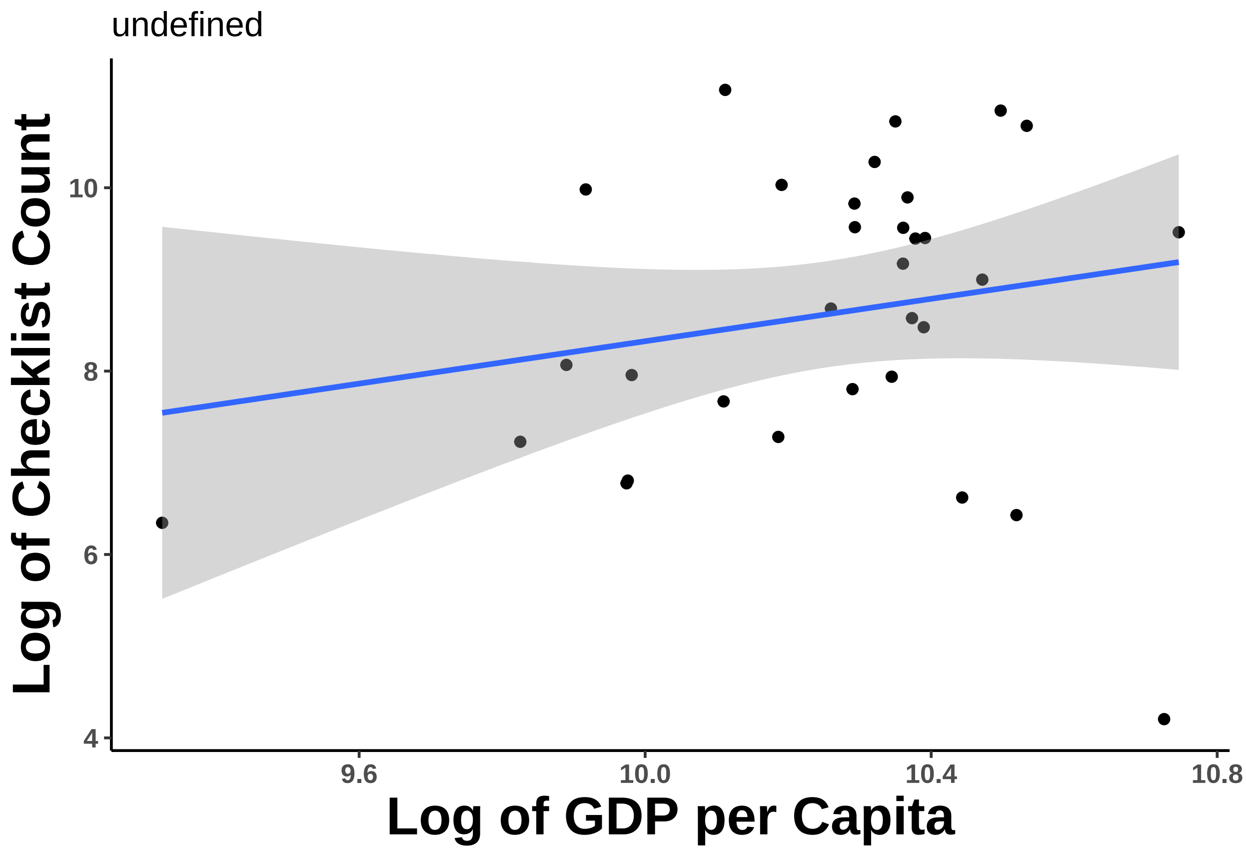
```
## 2865         Galax      Virginia           21769         106
## 1448       Neshoba   Mississippi           18693          62
## 1065      Magoffin      Kentucky           15118          32
## 1019    Crittenden      Kentucky           21375         112
## 2992         Boone West Virginia           21627         117
## 1082      Nicholas      Kentucky           20450          96
## 2143          Coal      Oklahoma           19752          89
## 1086        Owsley      Kentucky           13611          24
## 2607     Glasscock         Texas           31135         478
## 1108       Webster      Kentucky           20337         103
## 1008      Caldwell      Kentucky           20327         106
```

```
## Saving 6.5 x 4.5 in image
```



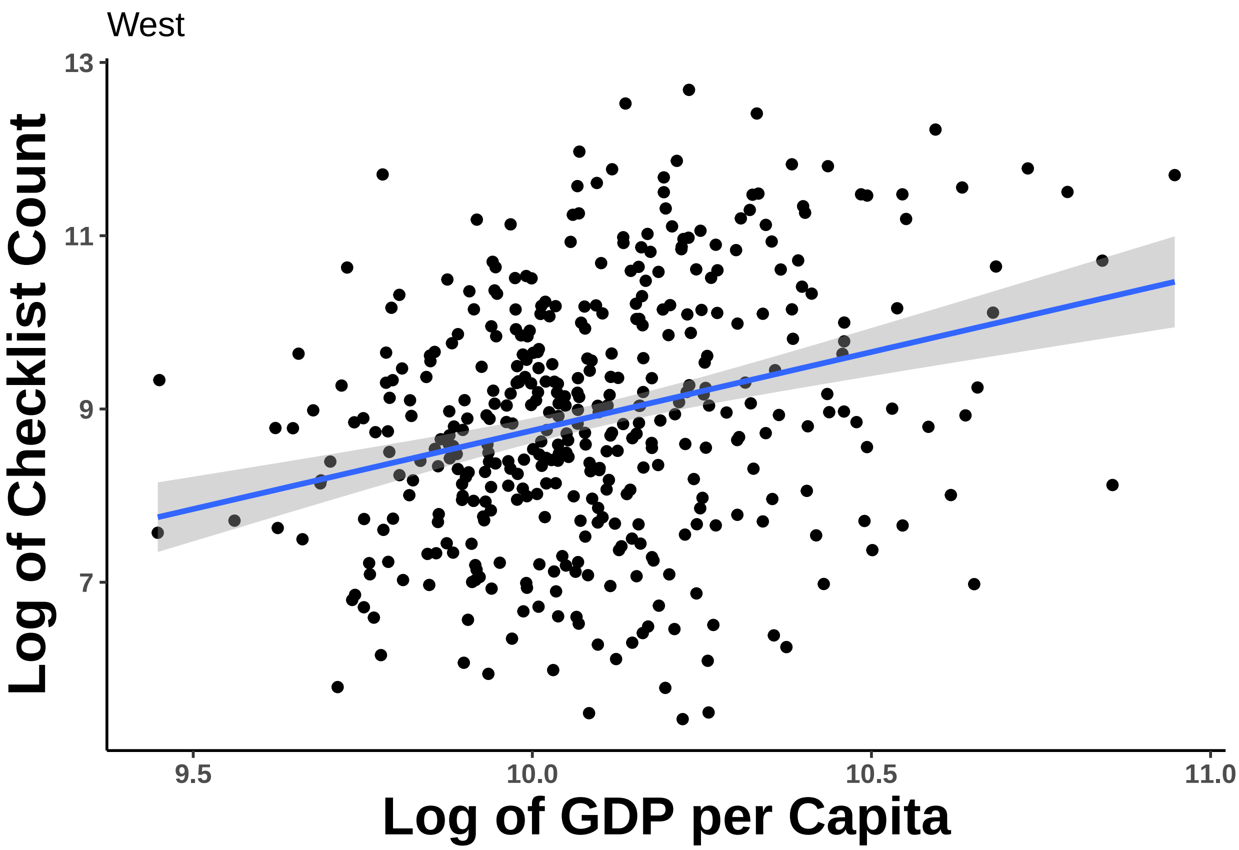

```
## [1] "Relationship: undefined y=(1.156x +/- 2.044) + (-3.23 +/- 20.98), r^2 = 0.04"
## [1] "Higher than Average Counties"
## [1] County          State           PerCapitaIncome eBird_count
## <0 rows> (or 0-length row.names)
## [1] "Lower than Average Counties"
##       County  State PerCapitaIncome eBird_count
## 547 Kalawao Hawaii           45515          67
```

```
## Saving 6.5 x 4.5 in image
```



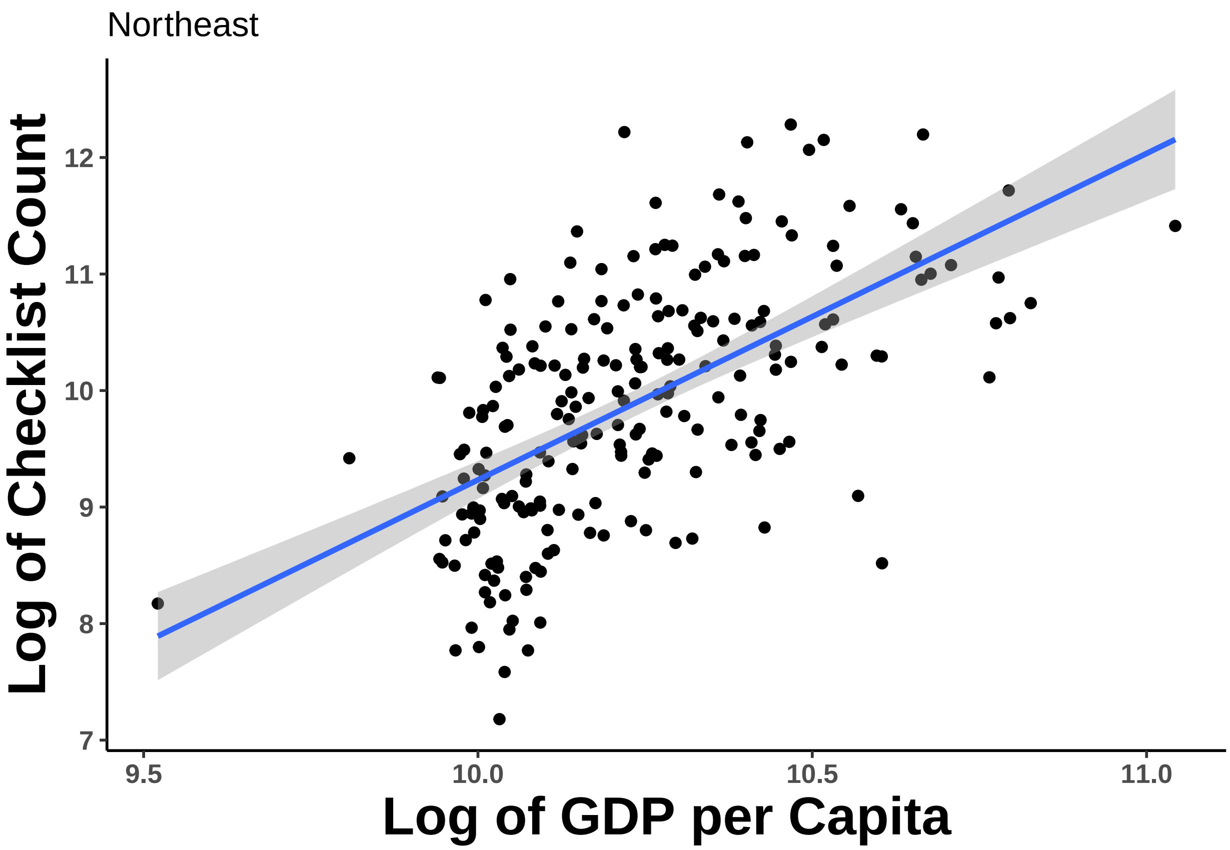

```
## [1] "Relationship: West y=(1.812x +/- 0.601) + (-9.37 +/- 6.06), r^2 = 0.08"
## [1] "Higher than Average Counties"
##           County      State PerCapitaIncome eBird_count
## 105         Pima    Arizona           25269      275358
## 203 Los Angeles California           27749      322379
## 107  Santa Cruz    Arizona           17664      121437
## 96       Cochise    Arizona           23608      158041
## 221   San Diego California           30668      245415
## 211    Monterey California           24775      128936
## 102    Maricopa    Arizona           27256      142069
## [1] "Lower than Average Counties"
##             County    State PerCapitaIncome eBird_count
## 1650        Wibaux  Montana           27492         226
## 1608        Fallon  Montana           28563         244
## 1647      Treasure  Montana           23948         242
## 3130      Niobrara  Wyoming           26797         324
## 1637      Richland  Montana           32036         519
## 1633 Powder River  Montana           28528         443
## 1605       Daniels  Montana           31449         594
## 283        Mineral Colorado           42255        1073
## 1601        Carter  Montana           24921         452
## 1630     Petroleum  Montana           22714         398
## 3139        Weston  Wyoming           28764         670
```

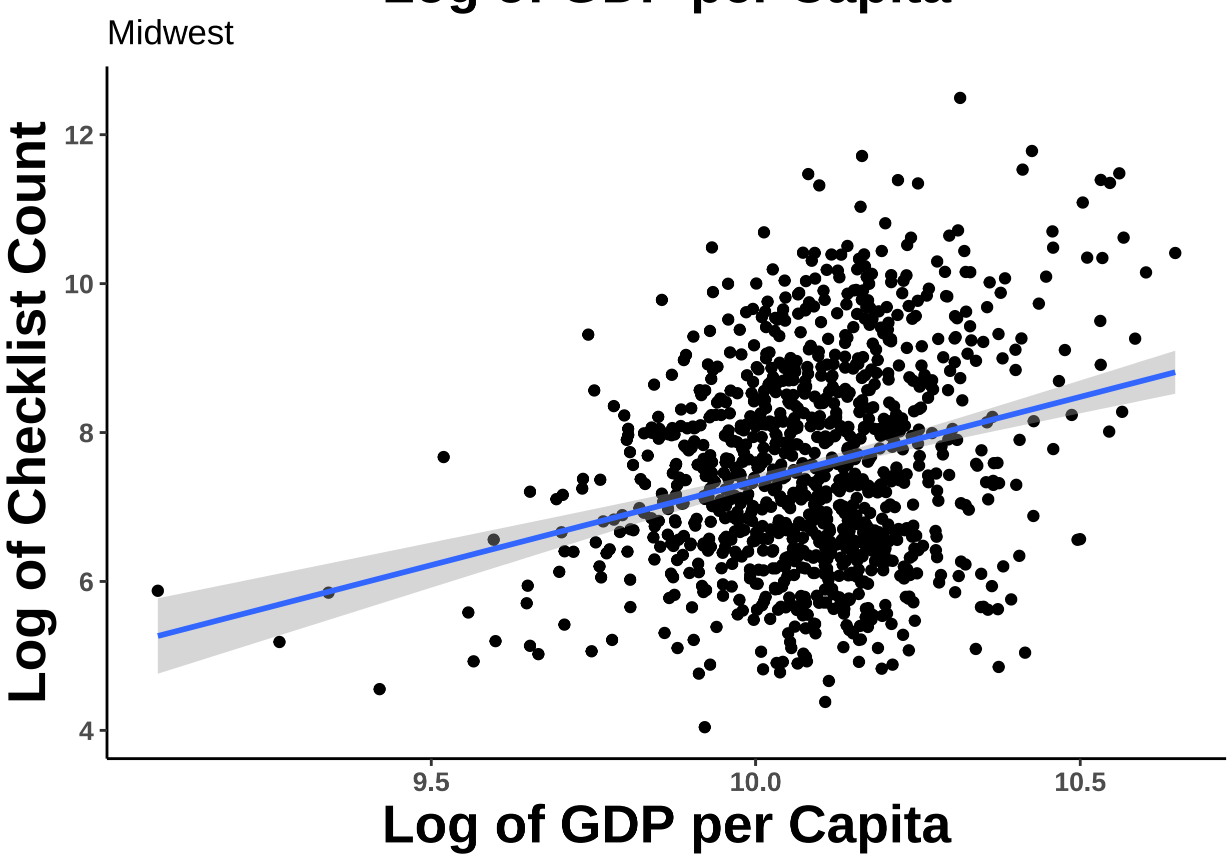
```
## Saving 6.5 x 4.5 in image
```




```
## [1] "Relationship: Northeast y=(2.801x +/- 0.515) + (-18.78 +/- 5.26), r^2 = 0.36"
## [1] "Higher than Average Counties"
##         County         State PerCapitaIncome eBird_count
## 1880 Tompkins      New York           27418      202458
## 1776 Cape May    New Jersey           32948      185427
## 1219     Essex Massachusetts          35167      215948
## 2255   Centre  Pennsylvania           25545       86291
## 1853   Monroe      New York           28732      110290
## [1] "Lower than Average Counties"
##         County        State PerCapitaIncome eBird_count
## 1865   Putnam     New York           40309        5003
## 2253  Cameron Pennsylvania           22747        1313
## 2298 Sullivan Pennsylvania           22923        1968
## 2309  Bristol Rhode Island           38893        8921
## 2265      Elk Pennsylvania           23738        2368
```

```
## Saving 6.5 x 4.5 in image
```






```
## [1] "Relationship: Midwest y=(2.26x +/- 0.505) + (-15.25 +/- 5.1), r^2 = 0.07"
```

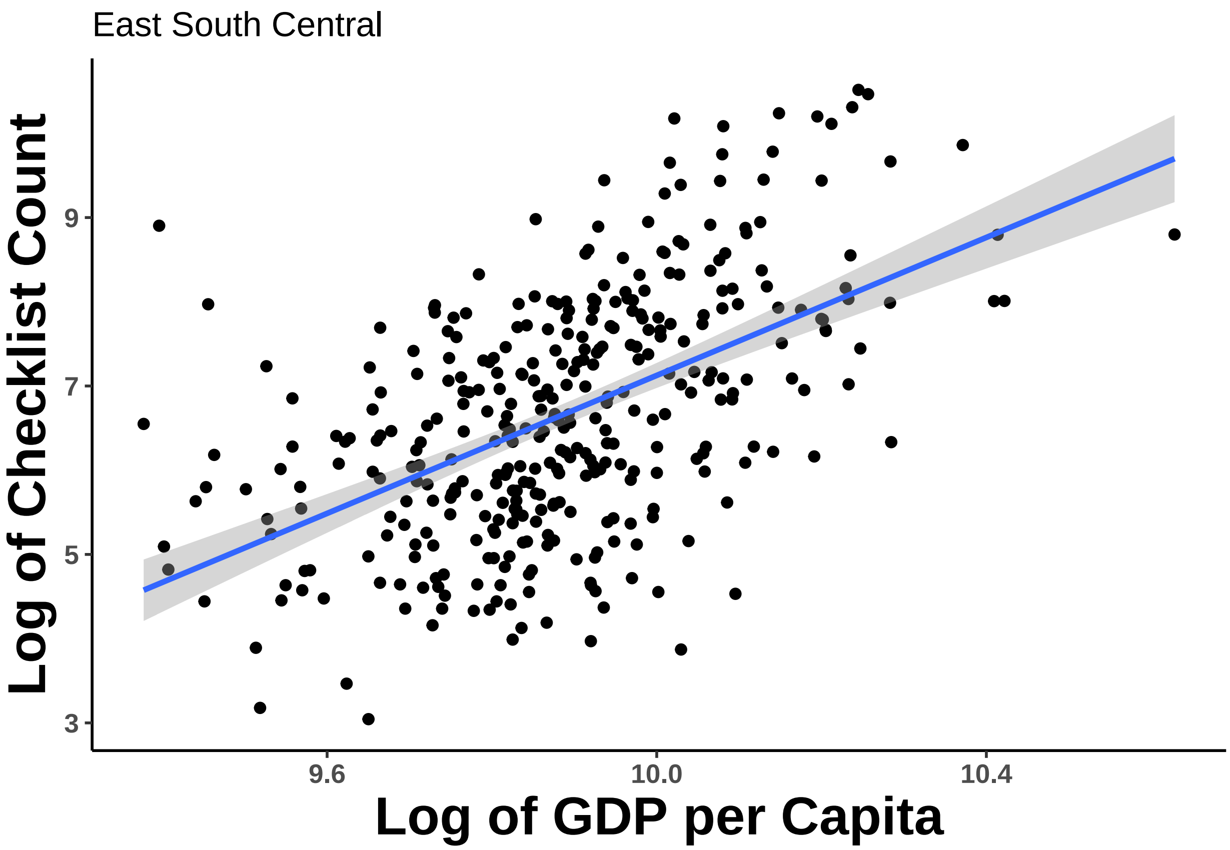
```
## [1] "Higher than Average Counties"
##           County     State PerCapitaIncome eBird_count
## 609         Cook  Illinois           30183      266739
## 1383  St. Louis Minnesota           25946      122312
## 2088       Lucas      Ohio           23885       95880
## 3085  Milwaukee Wisconsin           24295       82357
## 2058   Cuyahoga      Ohio           27423       88431
## 3057        Dane Wisconsin           33712      130748
## 2065   Franklin      Ohio           28283       84594
## 1269        Kent  Michigan           25889       61819
## 1310       Wayne  Michigan           22308       43869
## 1245   Chippewa  Michigan           20589       35800
## 1309  Washtenaw  Michigan           33231      101834
## 1490       Boone  Missouri           26895       49542
## 774  Tippecanoe   Indiana           23691       33359
## 615       DuPage  Illinois           38570       96845
## 744       Marion   Indiana           24124       33265
## 1338    Hennepin Minnesota           37485       88665
## 1298      Ottawa  Michigan           25371       36521
## 1266     Jackson  Michigan           22613       26664
## 1261      Ingham  Michigan           24754       32564
## 642         Lake  Illinois           38018       85150
## 1239     Berrien  Michigan           24013       29950
## 1265    Isabella  Michigan           19061       17701
## 715      Elkhart   Indiana           21109       21980
## 3060     Douglas Wisconsin           25129       32506
## 2102      Ottawa      Ohio           27979       40831
## 1289    Muskegon  Michigan           20621       19655
## [1] "Lower than Average Counties"
##         County        State PerCapitaIncome eBird_count
## 1997 Cavalier North Dakota           32028         128
## 2035   Towner North Dakota           33357         155
## 853  Mitchell         Iowa           24518          80
## 1574 Scotland     Missouri           20363          57
## 792   Audubon         Iowa           30919         163
## 833  Humboldt         Iowa           26746         125
## 2007   Griggs North Dakota           27197         132
## 859   Osceola         Iowa           24653         106
## 2011    Logan North Dakota           27887         160
## 2415  Sanborn South Dakota           25823         137
```

Second, by subregion:

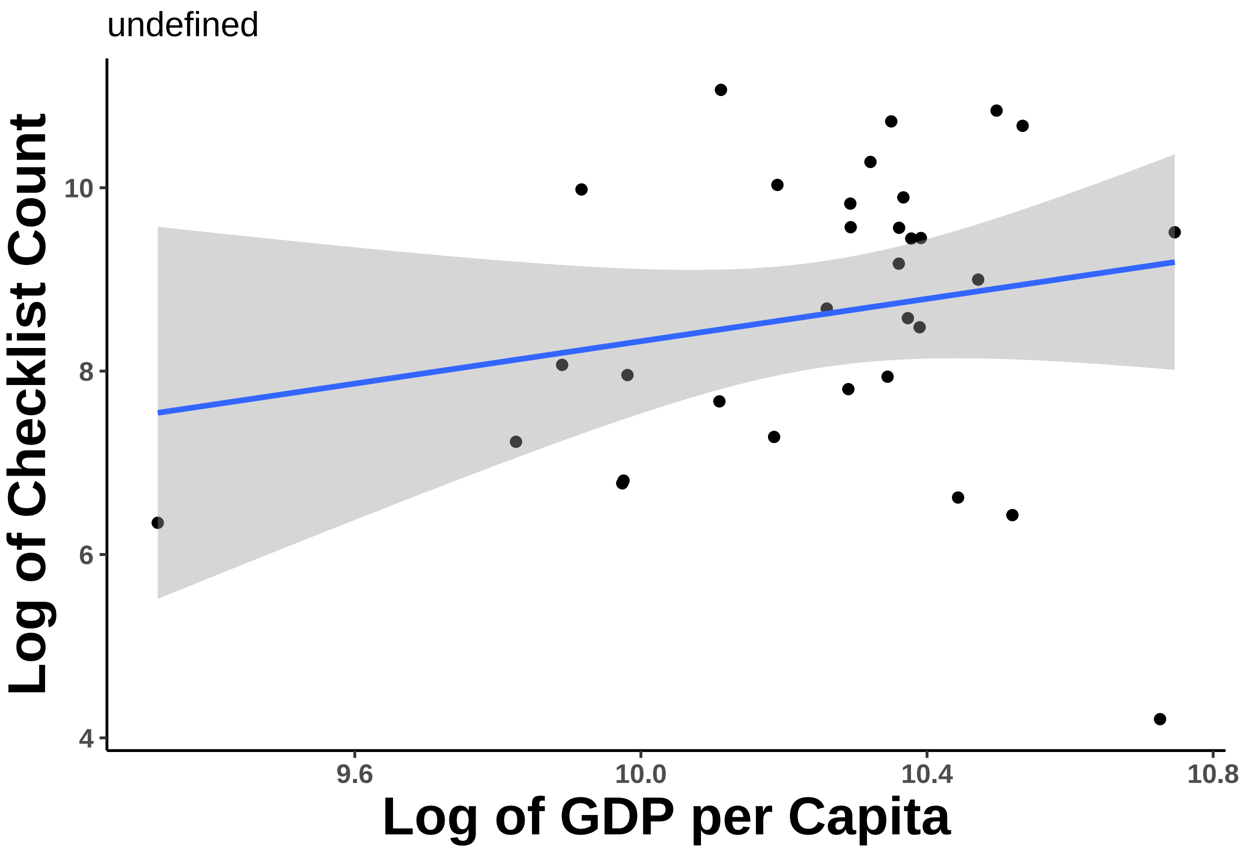
```
subregions=unique(x$Subregion)

for(i in 1:length(subregions)){

  subset.x=x[x$Subregion==subregions[i],]
  #subset.x=na.omit(subset.x)

  eq.x.1=lm(subset.x$logcheck~subset.x$logGDP)
  a.1.1=round(coef(eq.x.1)[1],2)#intercept
  b.1.1=round(coef(eq.x.1)[2],6)#slope
  r.x.1=round(summary(eq.x.1)$r.squared,2)
  a.err.1=round(summary(eq.x.1)$coefficients[3],2)
```

```r
  b.err.1=round(summary(eq.x.1)$coefficients[4],6)

  xval=min(subset.x$logGDP)+(max(subset.x$logGDP)-min(subset.x$logGDP))/2
  yval=1+max(subset.x$logcheck)

  a=ggplot(aes(x=logGDP,y=logcheck),data=subset.x)
  b=geom_point()
  c=geom_smooth(method=lm)
  d=theme_classic()
  d.1=theme(axis.title = element_text(face="bold",size=20),
            axis.text = element_text(size=10,face="bold"),
            legend.title = element_blank(),
            legend.text = element_text(size=14,face="bold"))
  w2=ggtitle(subregions[i])

  e=labs(x='Log of GDP per Capita',y='Log of Checklist Count')

  Fig3=a+b+c+d+e+d.1+w2
  ggsave(plot=Fig3,filename=paste0(filepath,"GDP-vs-Checklist_subregion_",subregions[i],".png"),dpi=400

  plot(Fig3)

  print(paste0('Relationship: ',subregions[i],' y=(',
               round(b.1.1,3),'x +/- ',
               round(2*b.err.1,3),') + (',round(a.1.1,3),
               " +/- ",round(2*a.err.1,3),')',', r^2 = ',round(r.x.1,2)))
}
```

```
## Saving 6.5 x 4.5 in image

## [1] "Relationship: East South Central y=(4.095x +/- 0.685) + (-33.83 +/- 6.76), r^2 = 0.28"

## Saving 6.5 x 4.5 in image
```

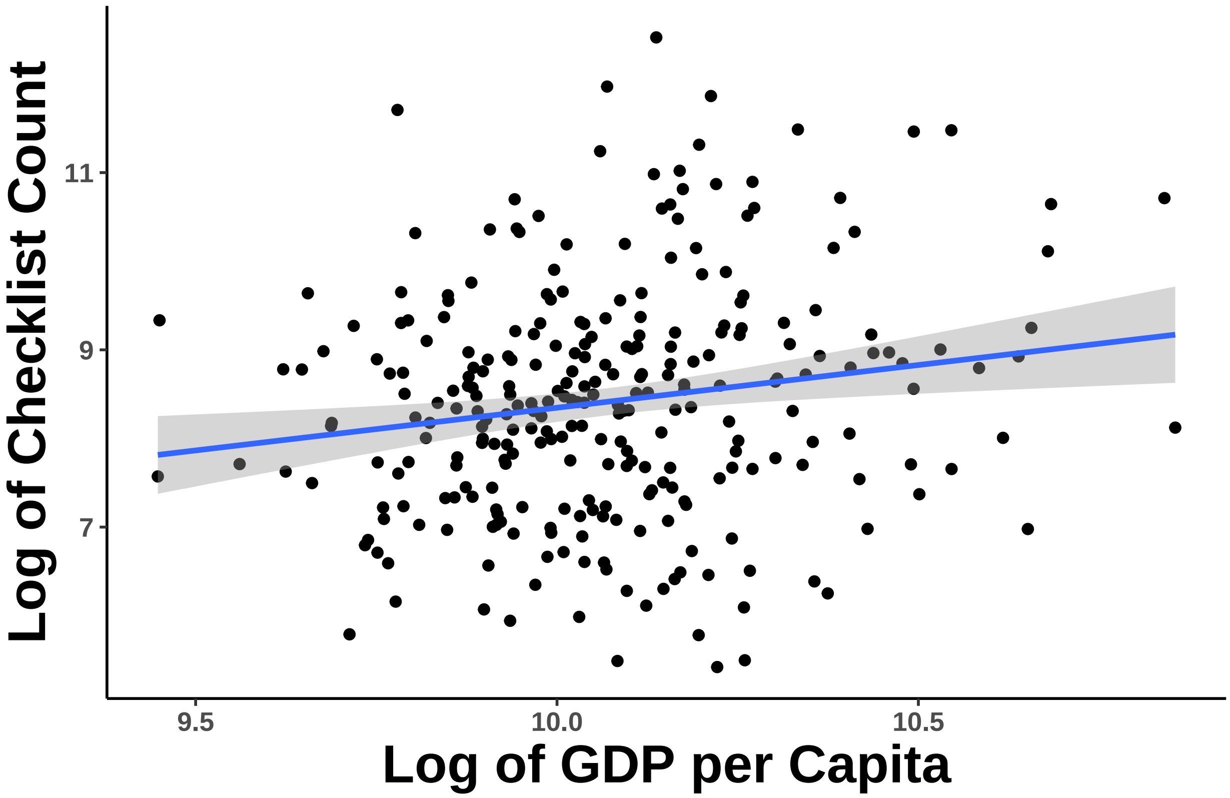



```
## [1] "Relationship: undefined y=(1.156x +/- 2.044) + (-3.23 +/- 20.98), r^2 = 0.04"
## Saving 6.5 x 4.5 in image
```

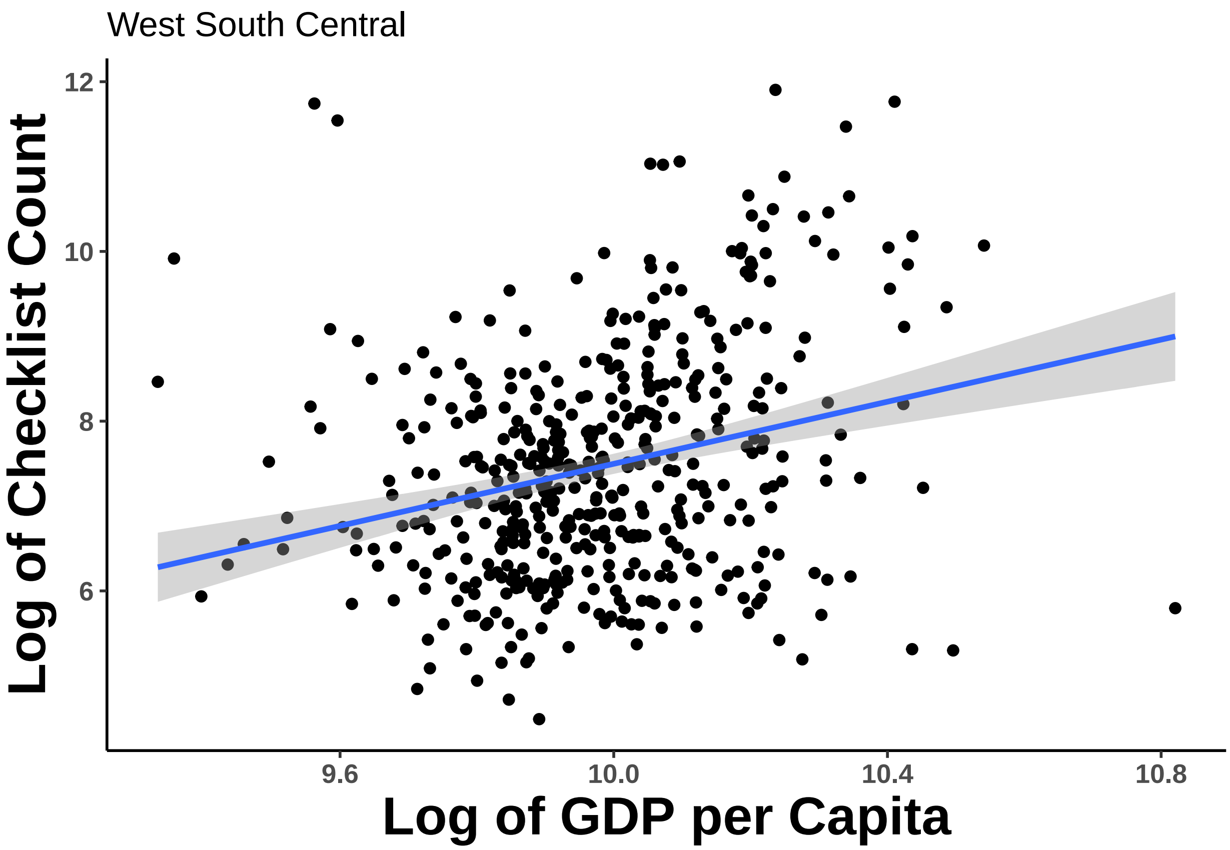



```
## [1] "Relationship: Mountain y=(0.964x +/- 0.673) + (-1.29 +/- 6.78), r^2 = 0.03"
## Saving 6.5 x 4.5 in image
```

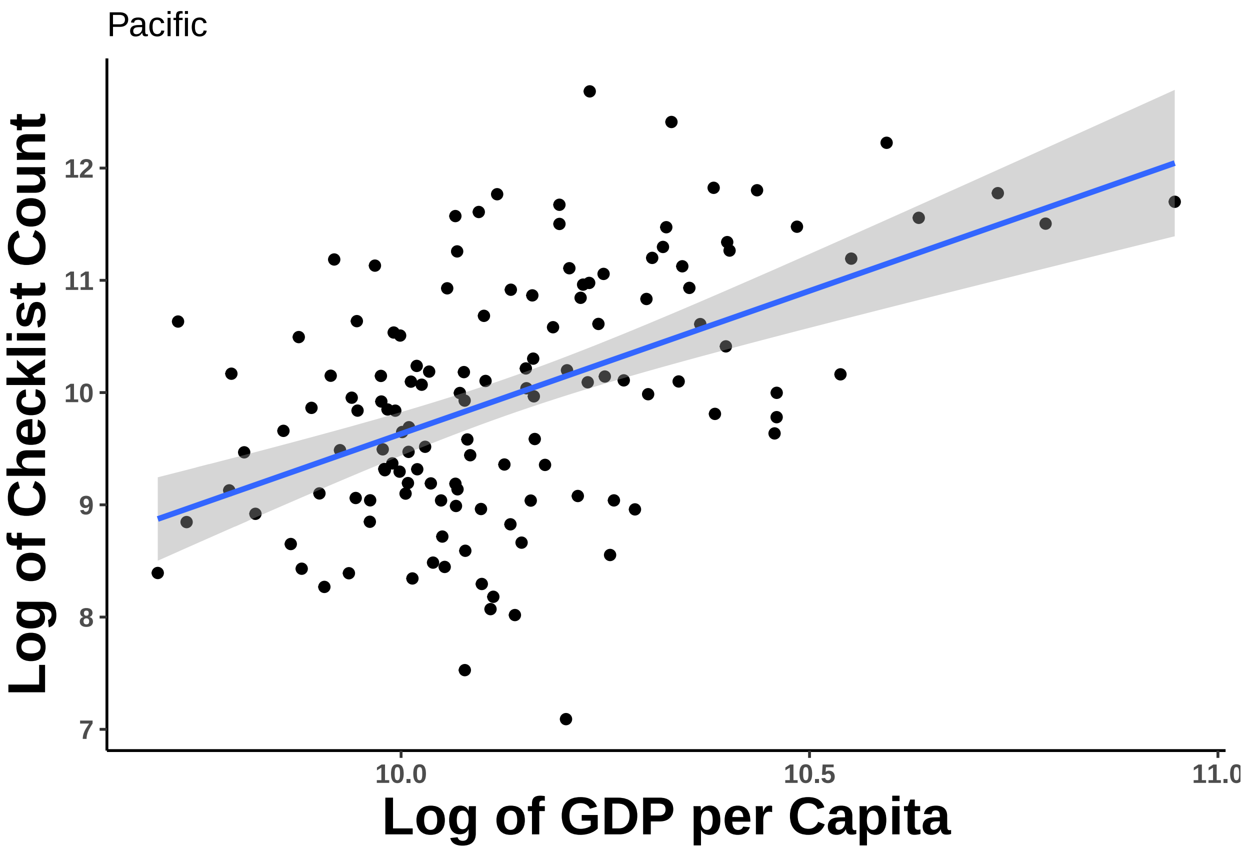
Mountain



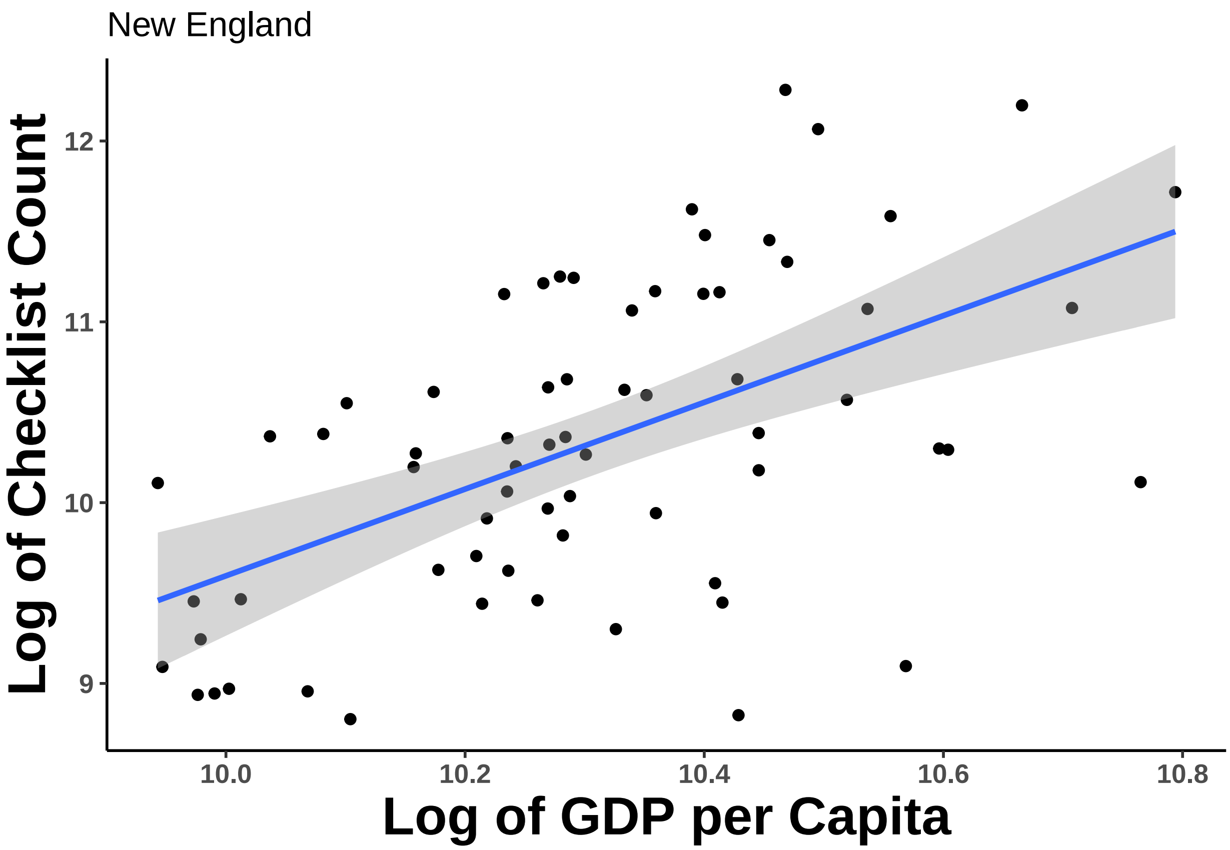

```
## [1] "Relationship: West South Central y=(1.829x +/- 0.616) + (-10.79 +/- 6.14), r^2 = 0.07"
## Saving 6.5 x 4.5 in image
```



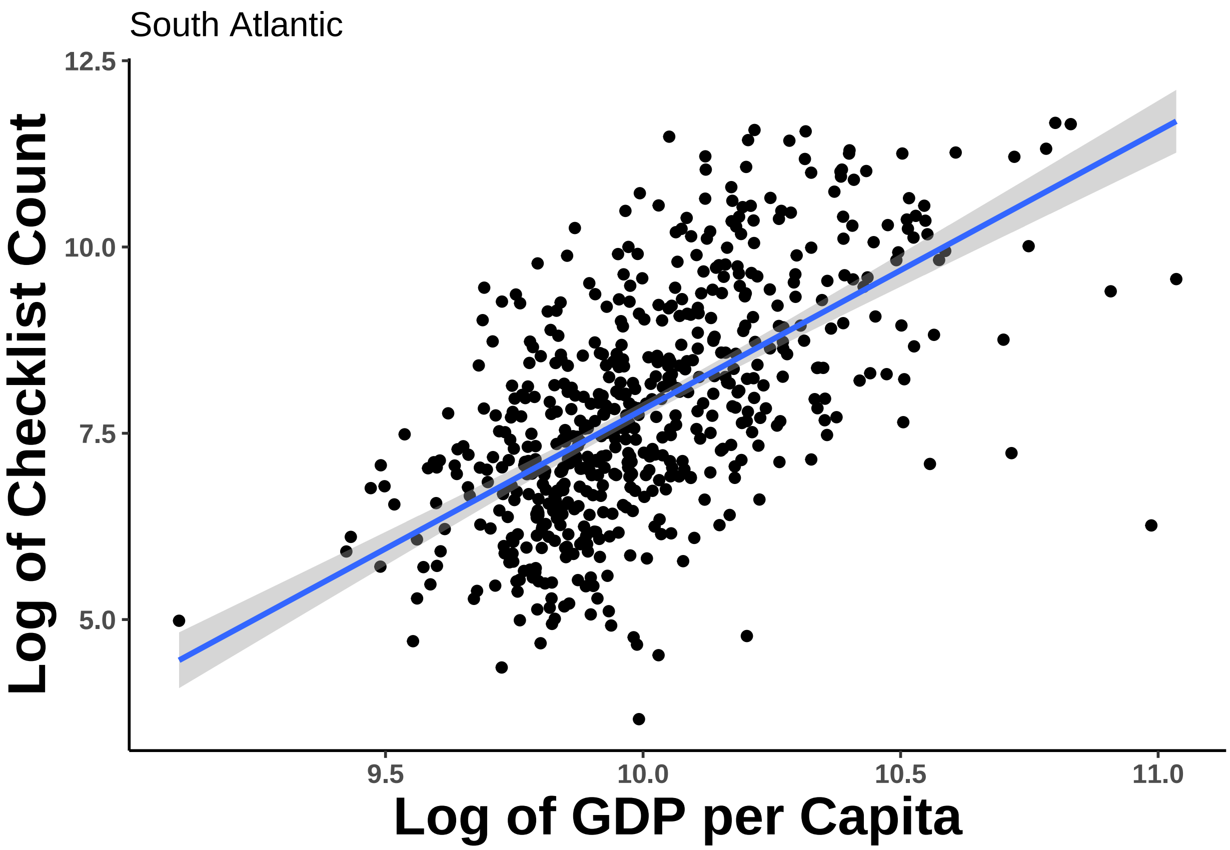

```
## [1] "Relationship: Pacific y=(2.547x +/- 0.783) + (-15.84 +/- 7.94), r^2 = 0.24"
## Saving 6.5 x 4.5 in image
```



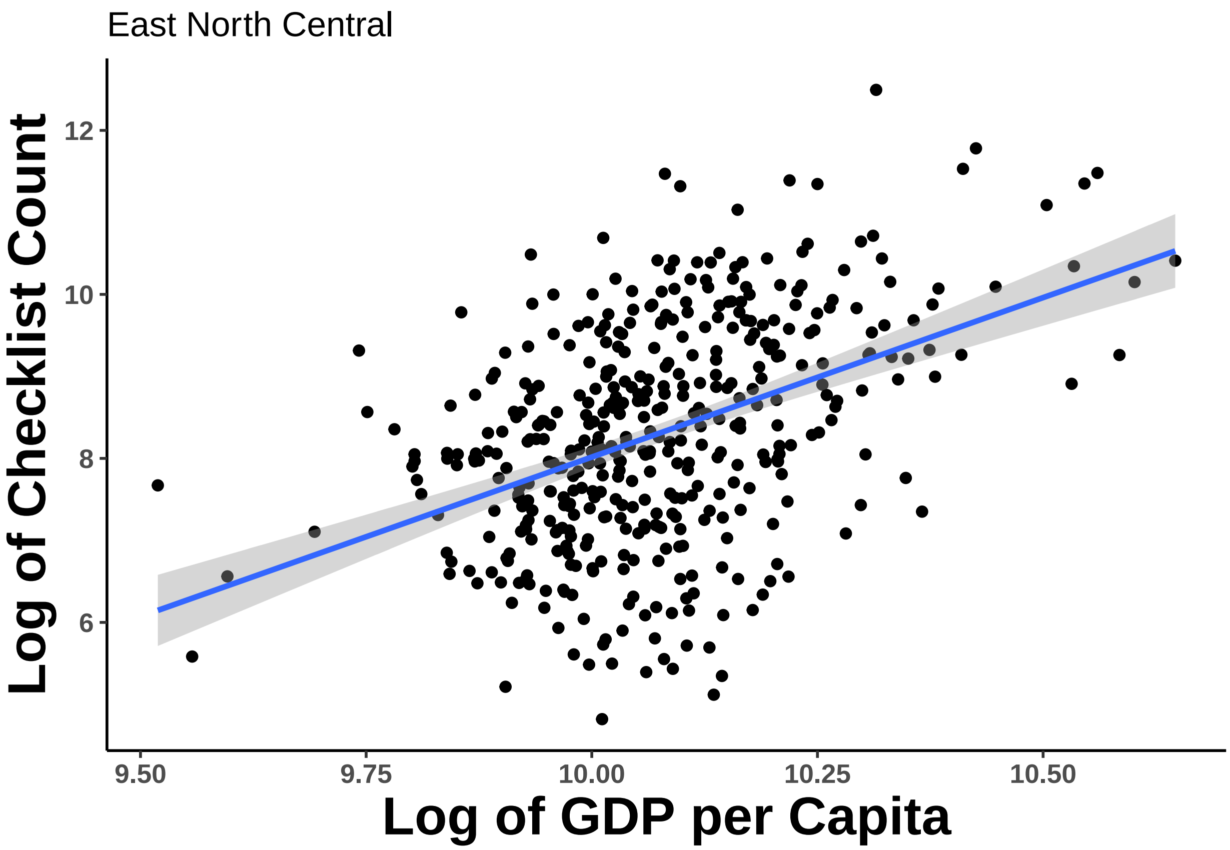

```
## [1] "Relationship: New England y=(2.398x +/- 0.91) + (-14.39 +/- 9.38), r^2 = 0.3"
## Saving 6.5 x 4.5 in image
```



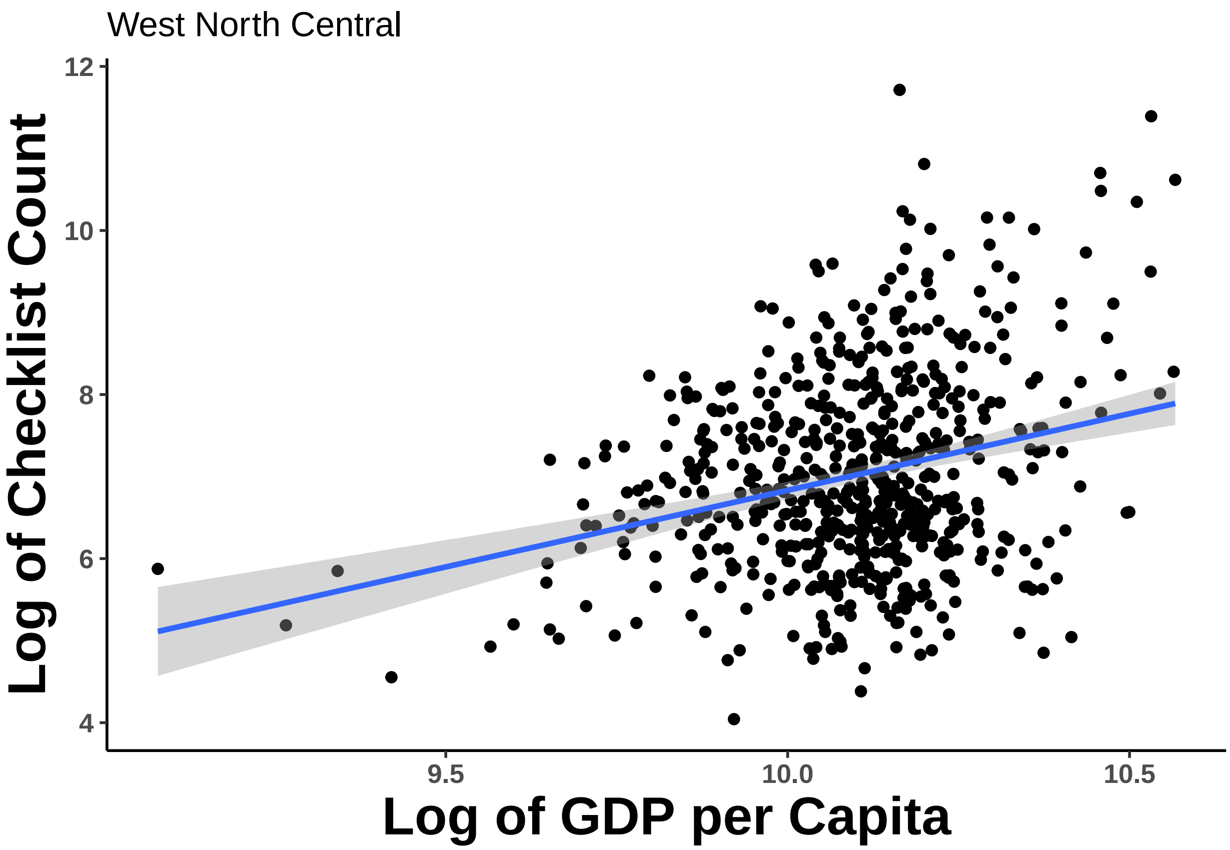

```
## [1] "Relationship: South Atlantic y=(3.737x +/- 0.405) + (-29.55 +/- 4.06), r^2 = 0.37"
## Saving 6.5 x 4.5 in image
```

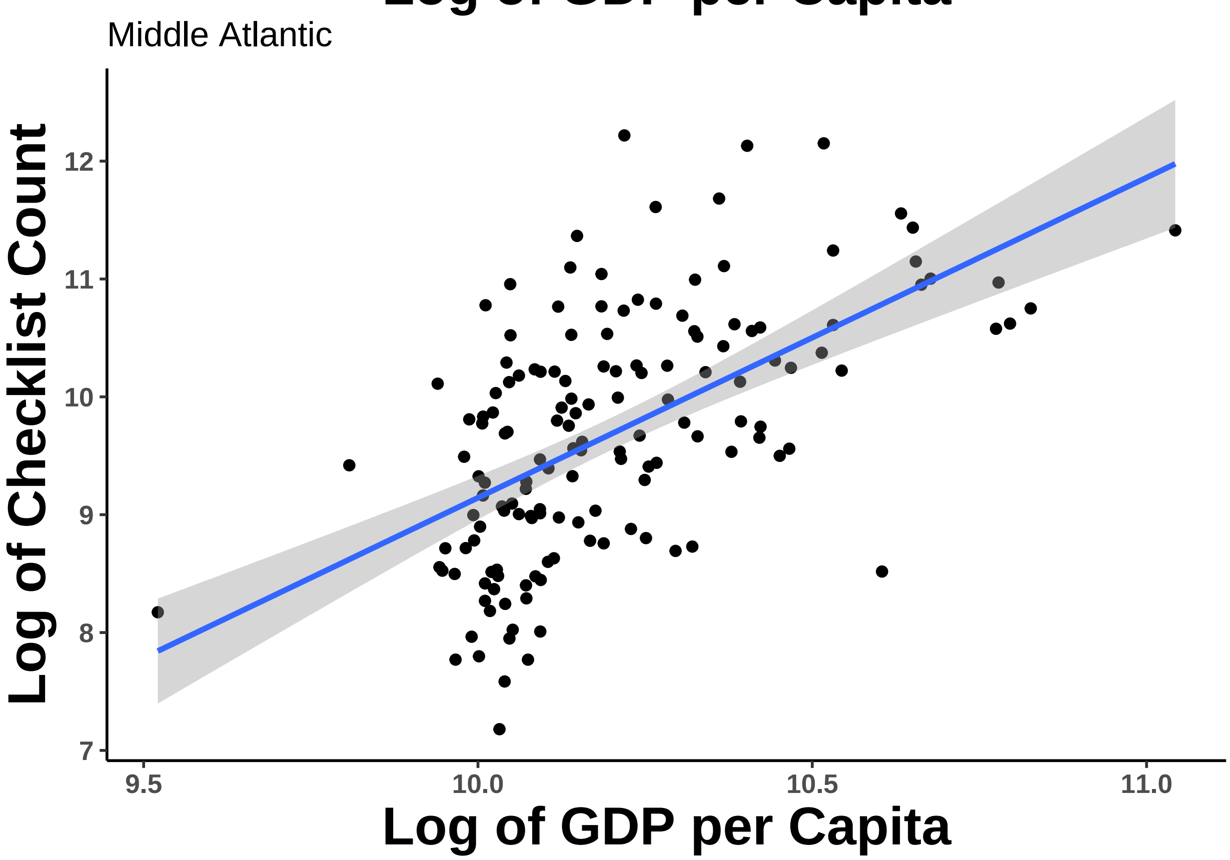



```
## [1] "Relationship: East North Central y=(3.889x +/- 0.77) + (-30.87 +/- 7.76), r^2 = 0.19"
## Saving 6.5 x 4.5 in image
```




```
## [1] "Relationship: West North Central y=(1.868x +/- 0.533) + (-11.85 +/- 5.38), r^2 = 0.07"
## Saving 6.5 x 4.5 in image
```






```
## [1] "Relationship: Middle Atlantic y=(2.717x +/- 0.63) + (-18.02 +/- 6.44), r^2 = 0.33"
```

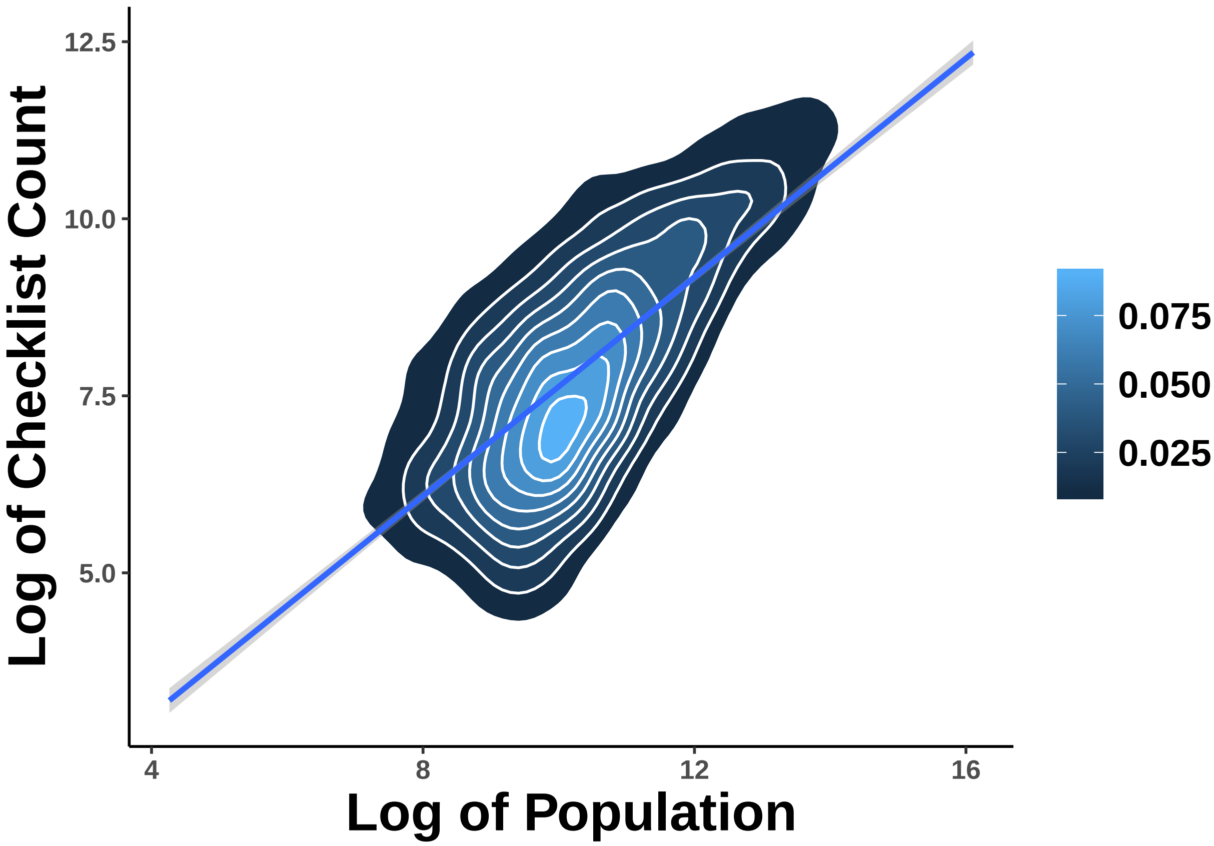
Interestingly, some areas show this connection a lot more strongly than others. This is likely related to more even birding effort in areas that are topographically heterogeneous and offering different bird communities in different regions. However, the South Atlantic has this diversity but still seems to strongly show birders focusing on wealthier counties.

Population Analyses

```
x$logpop=log(x$Population)

# get regression equation
eq.x.1=lm(x$logcheck~x$logpop)
a.1.1=round(coef(eq.x.1)[1],2)#intercept
b.1.1=round(coef(eq.x.1)[2],6)#slope
r.x.1=round(summary(eq.x.1)$r.squared,2)
a.err.1=round(summary(eq.x.1)$coefficients[3],2)
b.err.1=round(summary(eq.x.1)$coefficients[4],6)

# create plot
a=ggplot(aes(x=logpop,y=logcheck),data=x)
b=stat_density_2d(aes(fill=..level..),geom="polygon",colour="white")
c=geom_smooth(method=lm)
d=theme_classic()
d.1=theme(axis.title = element_text(face="bold",size=20),
          axis.text = element_text(size=10,face="bold"),
          legend.title = element_blank(),
          legend.text = element_text(size=14,face="bold"))

e=labs(x='Log of Population',y='Log of Checklist Count')
Fig3=a+b+c+d+e+d.1
ggsave(plot=Fig3,filename=paste0(filepath,"Pop-vs-Checklist_density.png"),dpi=400)
```

```
## Saving 6.5 x 4.5 in image
```

```
plot(Fig3)
```




```
paste0('Relationship: y=(',b.1.1,'x +/- ',2*b.err.1,') + (',
       a.1.1," +/- ",2*a.err.1,')',', r^2 = ',r.x.1)
```

```
## [1] "Relationship: y=(0.772586x +/- 0.028784) + (-0.1 +/- 0.3), r^2 = 0.48"
```

There is a fairly amorphous cloud of points, but the relationship is stronger than for GDP ($R^2 = 0.48$).

We can try to clip out extreme variation to see if this improves the fit.

It fits better, but not by a lot. There is still a lot of uncertainty in the cloud center. Let's look at the point density to determine what may be causing this.

Major regions

We are now going to repeat these analyses but with respect to population.

First, by major region:

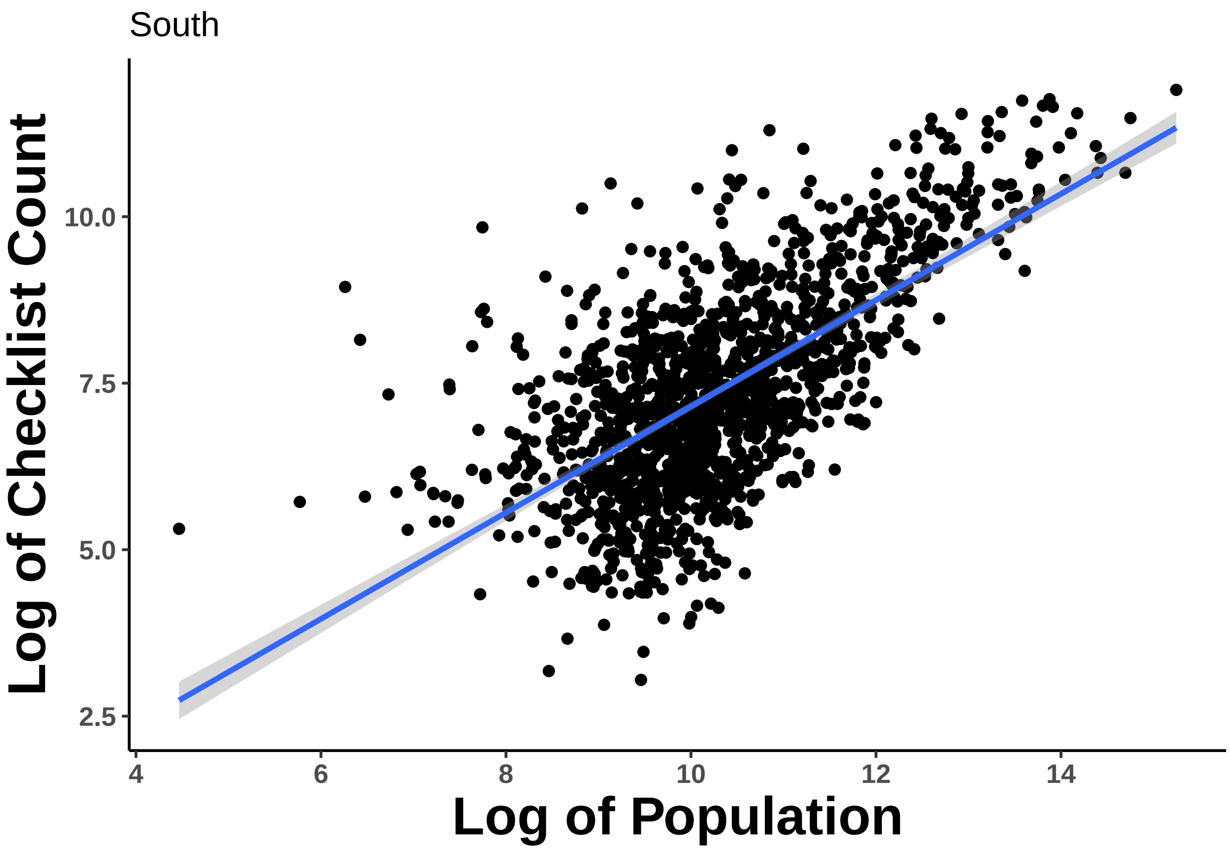
```
regions=unique(x$Region) #remove undefined

for(i in 1:length(regions)){
  subset.x=x[x$Region==regions[i],]
  #subset.x=na.omit(subset.x)

  eq.x.1=lm(subset.x$logcheck~subset.x$logpop)
  a.1.1=round(coef(eq.x.1)[1],2)#intercept
  b.1.1=round(coef(eq.x.1)[2],6)#slope
  r.x.1=round(summary(eq.x.1)$r.squared,2)
  a.err.1=round(summary(eq.x.1)$coefficients[3],2)
```

```
b.err.1=round(summary(eq.x.1)$coefficients[4],6)

xval=min(subset.x$logpop)+(max(subset.x$logpop)-min(subset.x$logpop))/2
yval=1+max(subset.x$logcheck)

a=ggplot(aes(x=logpop,y=logcheck),data=subset.x)
b=geom_point()
c=geom_smooth(method=lm)
d=theme_classic()
d.1=theme(axis.title = element_text(face="bold",size=20),
          axis.text = element_text(size=10,face="bold"),
          legend.title = element_blank(),
          legend.text = element_text(size=14,face="bold"))

e=labs(x='Log of Population',y='Log of Checklist Count')
w2=ggtitle(regions[i])

Fig3=a+b+c+d+e+d.1+w2
ggsave(plot=Fig3,filename=paste0(filepath,"Pop-vs-Checklist_region_",regions[i],".png"),dpi=400)

plot(Fig3)
print(paste0('Relationship: ',regions[i],' y=(',round(b.1.1,3),'x +/- ',round(2*b.err.1,3),
                               ') + (',round(a.1.1,3)," +/- ",round(2*a.err.1,3),')',
                               ', r^2 = ',round(r.x.1,2)))

#What are the biggest residual outliers?

dist2d=function(a1,a2,a3){
  v1=a2-a3
  v2=a1-a2
  m=cbind(v1,v2)
  d=det(m)/sqrt(sum(v1*v1))
  return(d)
}

a2=c(0,a.1.1)
val.test=(b.1.1*20)+a.1.1
a3=c(20,val.test)

subset.x$RDistance=0

for(j in 1:nrow(subset.x)){
  a1.x=subset.x$logpop[j]
  a1.y=subset.x$logcheck[j]
  a1=c(a1.x,a1.y)
  subset.x$RDistance[j]=dist2d(a1,a2,a3)
}

mu=mean(subset.x$RDistance)
sd.x=sd(subset.x$RDistance)

hi=qnorm(p=0.975,mean=mu,sd=sd.x)
lo=qnorm(p=0.025,mean=mu,sd=sd.x)
```

```r
  lows=subset.x[subset.x$RDistance<lo,]
  his=subset.x[subset.x$RDistance>hi,]

  lows=lows[order(lows$RDistance,decreasing=F),]
  his=his[order(his$RDistance,decreasing=T),]

  #Note that somehow hi and low are switched

  print("Higher than Average Counties")
  print(lows[,c("County","State","Population","eBird_count")])

  print("Lower than Average Counties")
  print(his[,c("County","State","Population","eBird_count")])
}
```

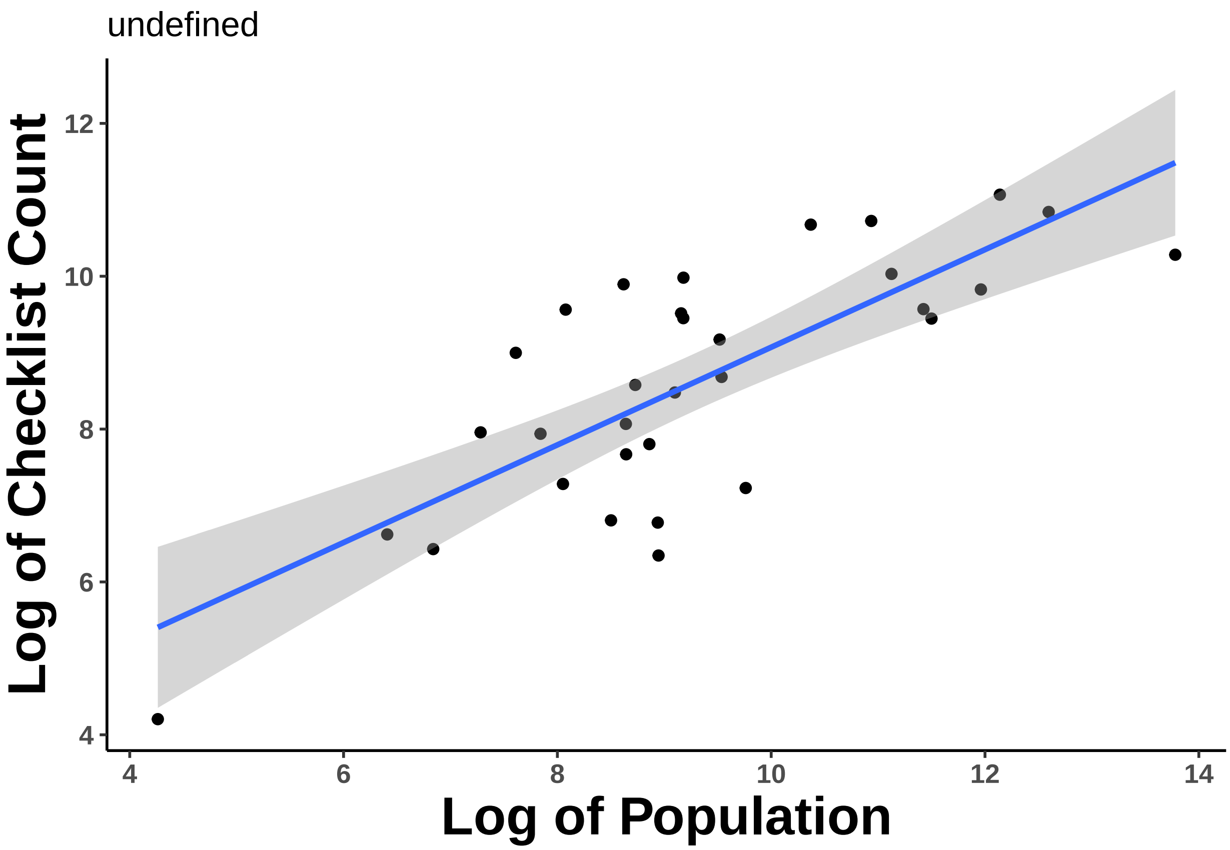
```
## Saving 6.5 x 4.5 in image

## [1] "Relationship: South y=(0.798x +/- 0.048) + (-0.83 +/- 0.5), r^2 = 0.44"
## [1] "Higher than Average Counties"
##             County          State Population eBird_count
## 2651        Kenedy          Texas        524        7669
## 2642    Jeff Davis          Texas       2311       18765
## 2542      Brewster          Texas       9244       36259
## 1123       Cameron      Louisiana       6789       24908
## 2682      McMullen          Texas        616        3466
## 2904   Northampton       Virginia      12339       26863
## 1915          Dare North Carolina      34289       59724
## 1214     Worcester       Maryland      51479       80641
## 2575     Culberson          Texas       2345        5520
## 2878      Highland       Virginia       2276        5257
## 2524       Aransas          Texas      23627       33641
## 2654        Kimble          Texas       4566        8954
## 2818      Accomack       Virginia      33289       38487
## 2141      Cimarron       Oklahoma       2432        4534
## 1211        Talbot       Maryland      37859       38327
## 2556      Chambers          Texas      35570       34880
## 362         Monroe        Florida      74213       61054
## 336       Franklin        Florida      11554       13546
## 1200    Dorchester       Maryland      32617       28993
## 1935          Hyde North Carolina       5771        7236
## 2589       Edwards          Texas       2070        3149
## 2742       Terrell          Texas        837        1527
## 1202       Garrett       Maryland      30014       24642
## 483       McIntosh        Georgia      14142       13093
## 2536        Blanco          Texas      10562        9446
## 2473          Lake      Tennessee       7773        7354
## 1208 Queen Anne's       Maryland      48166       31344
## 2671        Loving          Texas         87         203
## 2716       Refugio          Texas       7327        6766
## 2920       Radford       Virginia      16705       12759
## 2635     Hudspeth          Texas       3394        3539
## 383        Wakulla        Florida      30824       20069
## 1205          Kent       Maryland      20130       13975
## 3036        Tucker  West Virginia       7061        5917
```

```
## 2543      Briscoe          Texas       1615        1770
## 2713         Real          Texas       3348        3122
## 2664          Lee          Texas      16603       10897
## 448         Glynn        Georgia      80280       37626
## 2744 Throckmorton          Texas       1623        1653
## 2726     San Saba          Texas       6050        4649
## 2574       Crosby          Texas       6056        4399
## [1] "Lower than Average Counties"
##          County         State Population eBird_count
## 1068     Martin      Kentucky      12835          21
## 1065   Magoffin      Kentucky      13179          32
## 1448    Neshoba   Mississippi      29655          62
## 1017       Clay      Kentucky      21633          49
## 1468     Tippah   Mississippi      22080          54
## 1471      Union   Mississippi      27338          66
## 1438      Leake   Mississippi      23519          64
## 1027      Floyd      Kentucky      39448         104
## 1463      Smith   Mississippi      16414          53
## 1086     Owsley      Kentucky       4738          24
## 1088      Perry      Kentucky      28488         103
## 1457   Prentiss   Mississippi      25354         100
## 1052       Knox      Kentucky      31865         122
## 1037    Hancock      Kentucky       8608          48
## 1413     Copiah   Mississippi      29204         128
## 1066     Marion      Kentucky      19943          95
## 1051      Knott      Kentucky      16217          82
## 2992      Boone West Virginia      24517         117
## 3016   McDowell West Virginia      21651         111
## 3042       Wirt West Virginia       5796          39
## 1079     Morgan      Kentucky      13657          78
## 2479      Macon     Tennessee      22416         117
## 1401      Amite   Mississippi      13061          78
## 1449     Newton   Mississippi      21645         117
## 1437   Lawrence   Mississippi      12734          79
## 1078 Montgomery      Kentucky      26762         143
## 1405    Calhoun   Mississippi      14875          91
## 1004  Breathitt      Kentucky      13776          88
## 1475      Wayne   Mississippi      20675         123
## 1007     Butler      Kentucky      12746          85
## 1035      Green      Kentucky      11252          77

## Saving 6.5 x 4.5 in image
```




```
## [1] "Relationship: undefined y=(0.639x +/- 0.191) + (2.68 +/- 1.82), r^2 = 0.6"
## [1] "Higher than Average Counties"
## [1] County      State       Population  eBird_count
## <0 rows> (or 0-length row.names)
## [1] "Lower than Average Counties"
##      County  State Population eBird_count
## 81 Kusilvak Alaska       7678         570
```

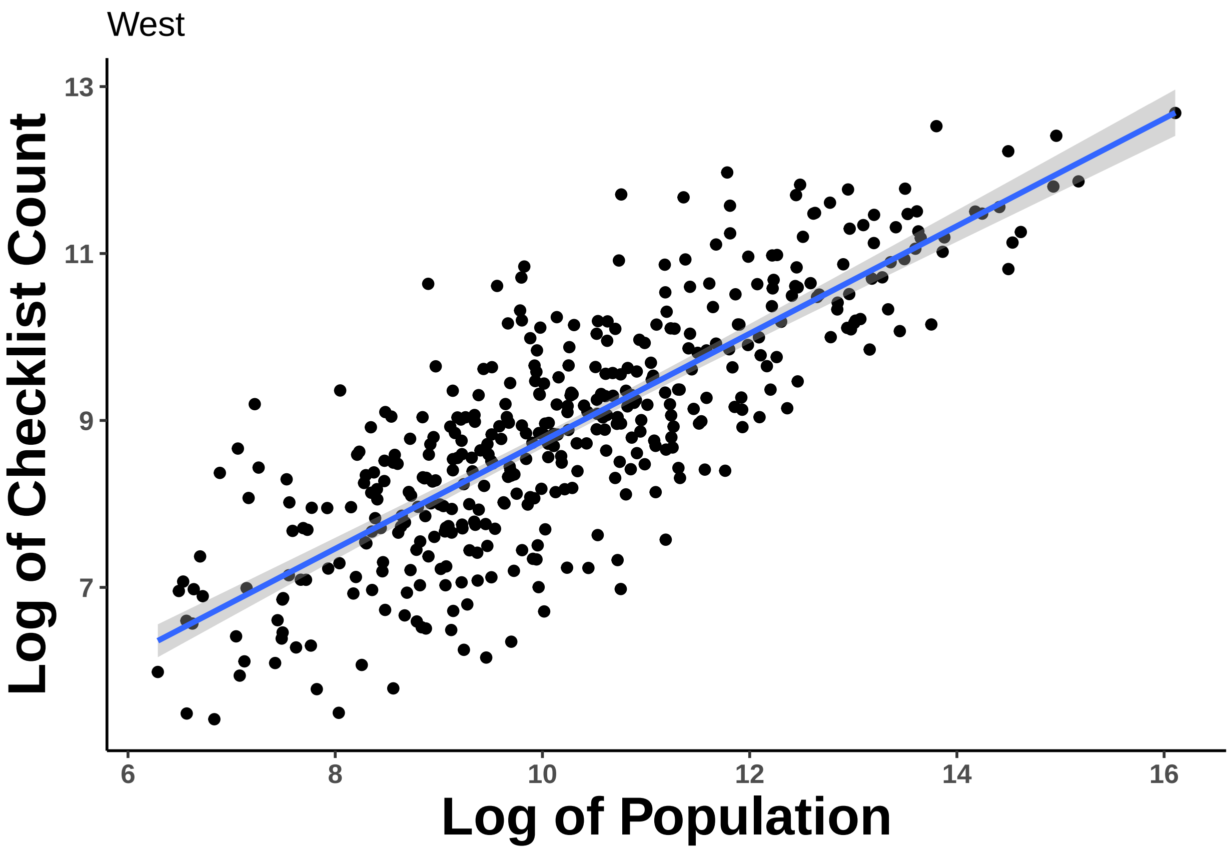
```
## Saving 6.5 x 4.5 in image
```




```
## [1] "Relationship: West y=(0.644x +/- 0.046) + (2.31 +/- 0.48), r^2 = 0.66"
## [1] "Higher than Average Counties"
##           County      State Population eBird_count
## 2218      Harney     Oregon       7314       41628
## 107   Santa Cruz    Arizona      47122      121437
## 272      Jackson   Colorado       1371        9851
## 198         Inyo California      18482       51238
## 210         Mono California      14217       40603
## 1808 Los Alamos New Mexico      17979       44896
## 96       Cochise    Arizona     131038      158041
## 2207      Benton     Oregon      85989      117380
## 230       Sierra California       3127       11605
## 186       Alpine California       1165        5786
## 1821     Socorro New Mexico      17756       30255
## 2226     Lincoln     Oregon      46070       55066
## [1] "Lower than Average Counties"
##           County      State Population eBird_count
## 3119    Campbell    Wyoming      46901        1075
## 570     Franklin      Idaho      12801         473
## 1616        Hill    Montana      16301         572
## 576       Jerome      Idaho      22391         822
## 581      Lincoln      Idaho       5221         327
## 1637    Richland    Montana      10318         519
## 1608      Fallon    Montana       3085         244
## 1810    McKinley New Mexico      72373        1942
## 555      Bingham      Idaho      45485        1521
## 1642 Silver Bow     Montana      34322        1384
```

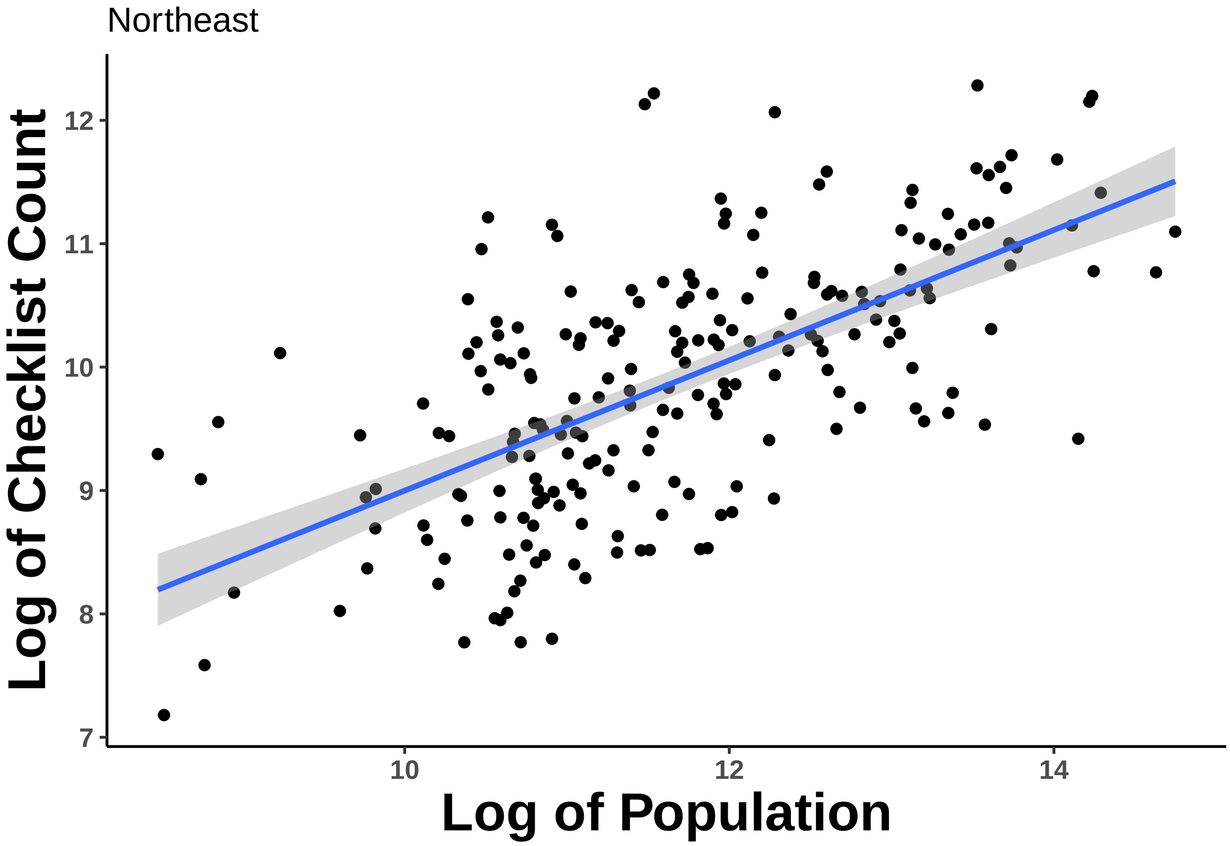
```
## 2778      Carbon        Utah       21227          1100
## 1606      Dawson     Montana        9132           658
## 2794     Sanpete        Utah       27930          1388
```

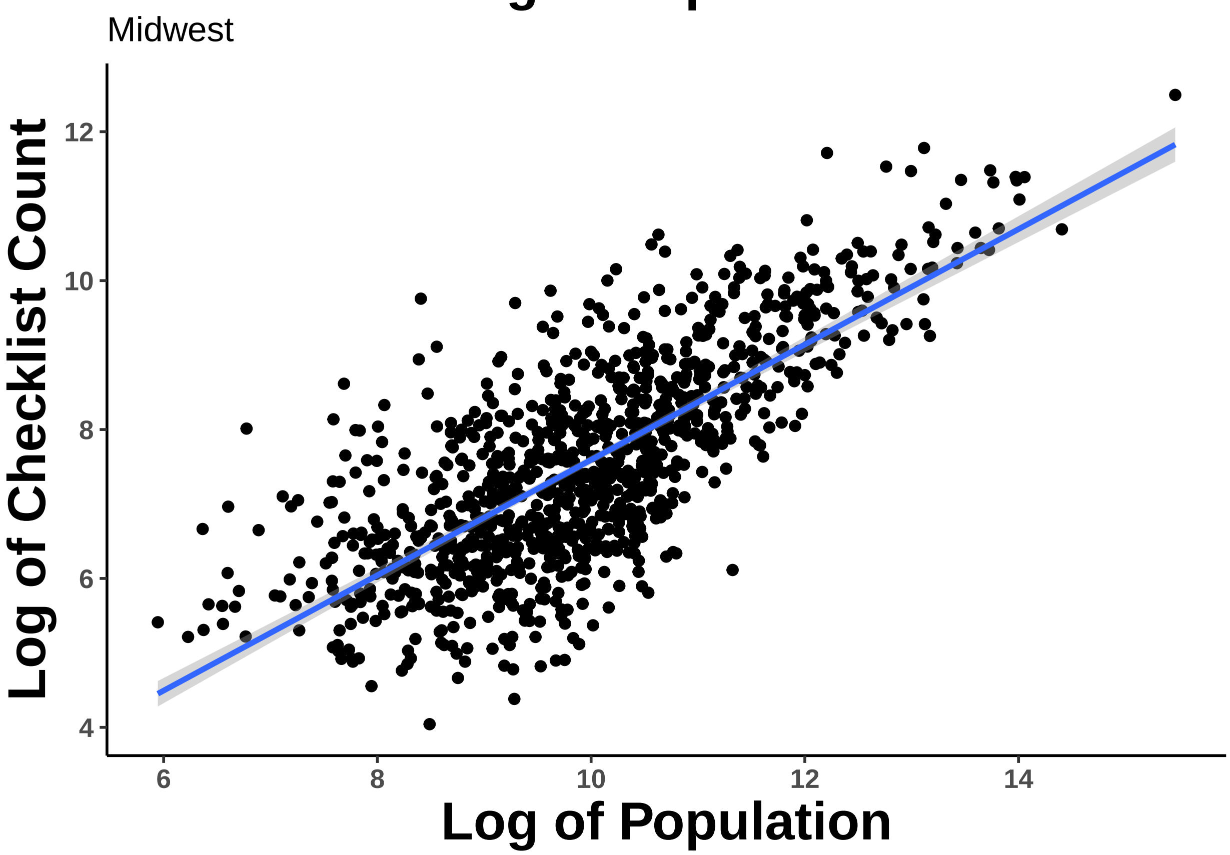
```
## Saving 6.5 x 4.5 in image
```




```
## [1] "Relationship: Northeast y=(0.528x +/- 0.086) + (3.72 +/- 1.02), r^2 = 0.41"
## [1] "Higher than Average Counties"
##          County         State Population eBird_count
## 1880   Tompkins      New York     102270      202458
## 1776   Cape May    New Jersey      96684      185427
## 2804     Addison       Vermont      36811       74105
## 1215 Barnstable Massachusetts     215449      173792
## 1874      Seneca      New York      35359       57302
## 1180     Hancock         Maine      54557       69801
## 2817     Windsor       Vermont      56416       63758
## [1] "Lower than Average Counties"
##         County        State Population eBird_count
## 1828      Bronx     New York    1397315       12323
## 2302   Venango Pennsylvania      54590        2437
## 2274 Jefferson Pennsylvania      45015        2369
```

```
## Saving 6.5 x 4.5 in image
```






```
## [1] "Relationship: Midwest y=(0.775x +/- 0.041) + (-0.15 +/- 0.42), r^2 = 0.57"
```

```
## [1] "Higher than Average Counties"
##         County        State Population eBird_count
## 3063  Florence    Wisconsin       4477       17287
## 1991  Billings North Dakota        876        3019
## 1270  Keweenaw     Michigan       2181        5514
## 1349      Lake    Minnesota      10825       16304
## 1327      Cook    Minnesota       5195        9063
## 979   Stafford       Kansas       4391        7647
## 3048  Bayfield    Wisconsin      15071       19233
## 2102    Ottawa         Ohio      41372       40831
## 1245  Chippewa     Michigan      38760       35800
## 2033    Steele North Dakota       1976        3420
## 1383 St. Louis    Minnesota     200327      122312
## 3059      Door    Wisconsin      27826       25673
## 1263     Iosco     Michigan      25662       22054
## 3060   Douglas    Wisconsin      43994       32506
## 951     Morton       Kansas       3182        4143
## 3046   Ashland    Wisconsin      16068       13596
## 1248  Crawford     Michigan      14017       11850
## 2009    Kidder North Dakota       2429        2958
## 1273  Leelanau     Michigan      21682       16057
## 1524      Holt     Missouri       4771        4829
## 935      Kiowa       Kansas       2527        2940
## 1230     Alger     Michigan       9497        7890
## 2031     Slope North Dakota        738        1058
## 3065    Forest    Wisconsin       9255        7444
## 2380   Douglas South Dakota       3000        3101
## 3051   Burnett    Wisconsin      15449       10899
## 3069      Iowa    Wisconsin      23709       15176
## 1709      Loup     Nebraska        581         784
## 3108     Vilas    Wisconsin      21401       12668
## 1279  Manistee     Michigan      24626       13898
## 980    Stanton       Kansas       2210        2104
## [1] "Lower than Average Counties"
##           County    State Population eBird_count
## 853     Mitchell     Iowa      10739          80
## 729       Howard  Indiana      82795         452
## 811     Crawford     Iowa      17205         135
## 860         Page     Iowa      15838         134
## 606         Clay Illinois      13744         124
## 610     Crawford Illinois      19707         167
## 1574    Scotland Missouri       4859          57
## 1595      Wright Missouri      18643         181
## 856   Montgomery     Iowa      10625         119
## 849      Mahaska     Iowa      22428         215
## 881      Webster     Iowa      37626         333
## 833     Humboldt     Iowa       9776         125
## 763     Randolph  Indiana      25975         273
## 877      Wapello     Iowa      35469         363
## 765         Rush  Indiana      17257         220
## 886       Wright     Iowa      13092         184
## 859      Osceola     Iowa       6335         106
## 2121    Van Wert     Ohio      28685         365
## 1520      Grundy Missouri      10282         165
```

```
## 689        Wayne Illinois      16674         244
## 810      Clinton     Iowa      48896         565
## 798  Buena Vista     Iowa      20350         287
## 2109      Putnam     Ohio      34339         441
## 767       Shelby  Indiana      44511         541
## 815     Delaware     Iowa      17665         265
## 649     Macoupin Illinois      47462         575
## 813        Davis     Iowa       8755         157
```

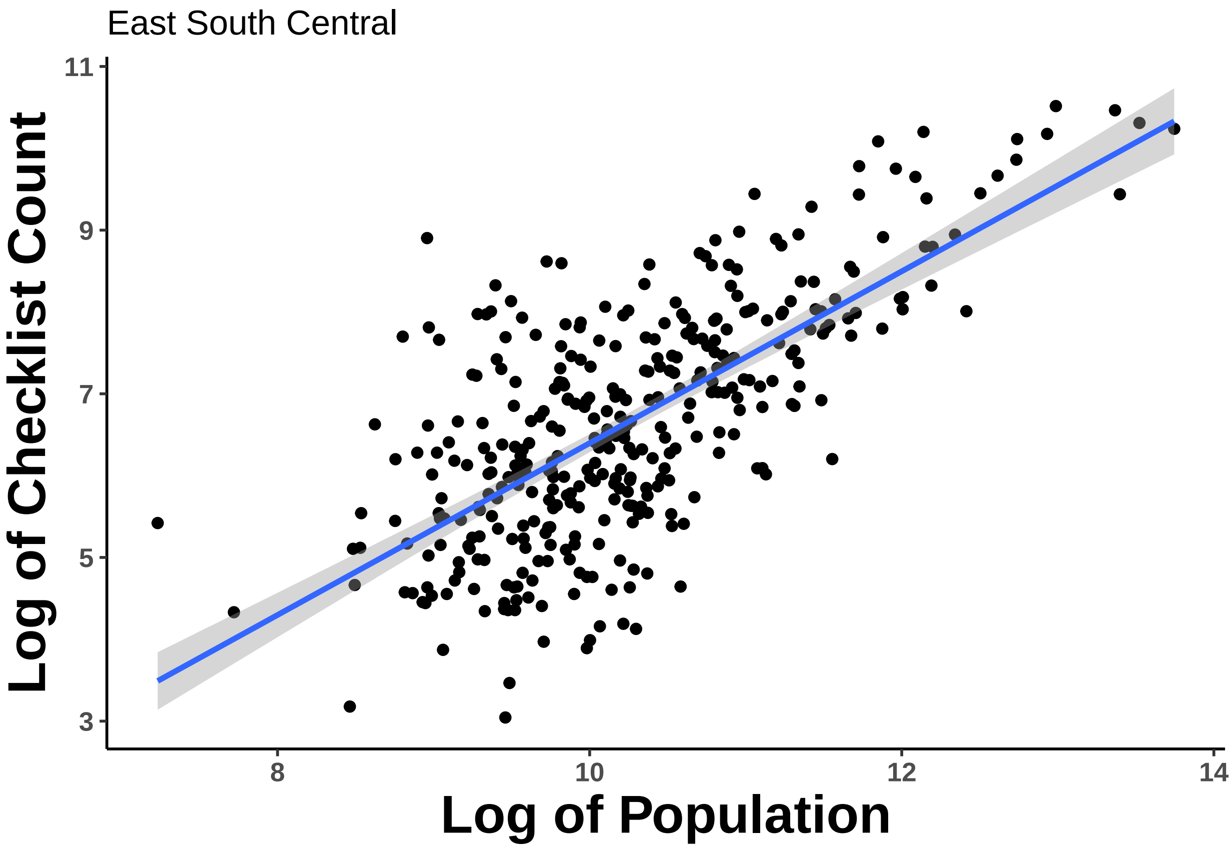
Regional Analyses

Second, by subregion:

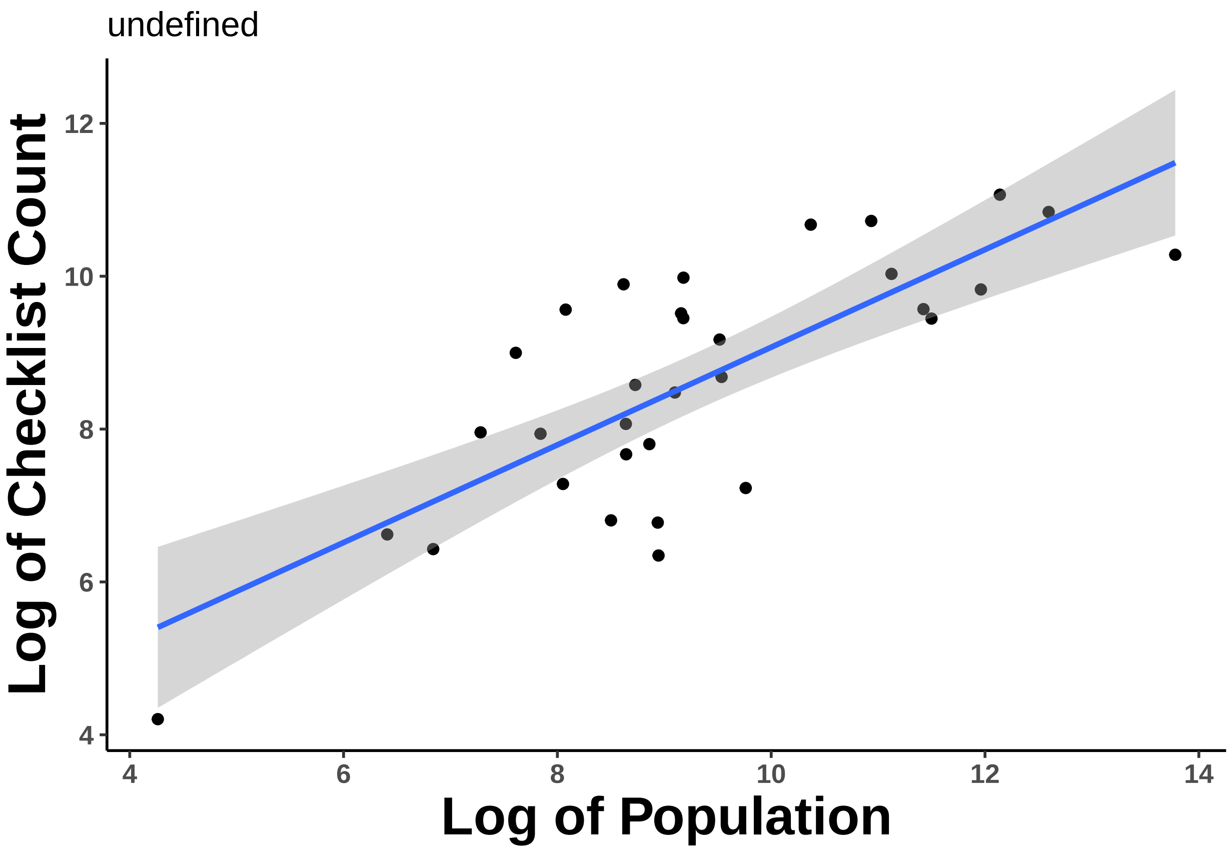
```
subregions=unique(x$Subregion)

for(i in 1:length(subregions)){

  subset.x=x[x$Subregion==subregions[i],]
  #subset.x=na.omit(subset.x)

  eq.x.1=lm(subset.x$logcheck~subset.x$logpop)
  a.1.1=round(coef(eq.x.1)[1],2)#intercept
  b.1.1=round(coef(eq.x.1)[2],6)#slope
  r.x.1=round(summary(eq.x.1)$r.squared,2)
  a.err.1=round(summary(eq.x.1)$coefficients[3],2)
  b.err.1=round(summary(eq.x.1)$coefficients[4],6)

  xval=min(subset.x$logGDP)+(max(subset.x$logpop)-min(subset.x$logpop))/2
  yval=1+max(subset.x$logcheck)

  a=ggplot(aes(x=logpop,y=logcheck),data=subset.x)
  b=geom_point()
  c=geom_smooth(method=lm)
  d=theme_classic()
  d.1=theme(axis.title = element_text(face="bold",size=20),
            axis.text = element_text(size=10,face="bold"),
            legend.title = element_blank(),
            legend.text = element_text(size=14,face="bold"))
  w2=ggtitle(subregions[i])

  e=labs(x='Log of Population',y='Log of Checklist Count')

  Fig3=a+b+c+d+e+d.1+w2
  ggsave(plot=Fig3,filename=paste0(filepath,"Pop-vs-Checklist_region_",subregions[i],".png"),dpi=400)

  plot(Fig3)
  print(paste0('Relationship ', subregions[i],': y=(',
               round(b.1.1,3),'x +/- ',
               round(2*b.err.1,3),') + (',round(a.1.1,3),
               " +/- ",round(2*a.err.1,3),')',', r^2 = ',round(r.x.1,2)))
}
```

```
## Saving 6.5 x 4.5 in image
```

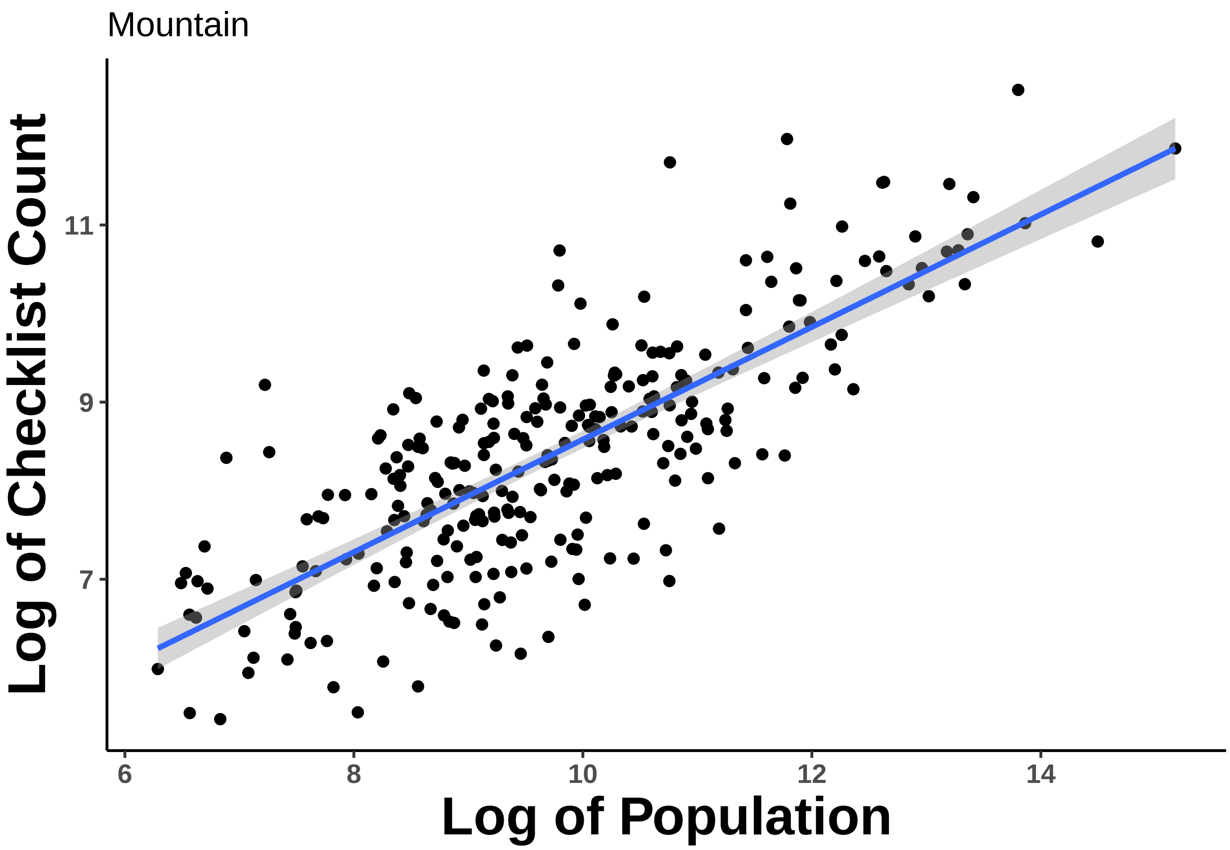
```
## [1] "Relationship East South Central: y=(1.05x +/- 0.113) + (-4.1 +/- 1.16), r^2 = 0.49"
## Saving 6.5 x 4.5 in image
```



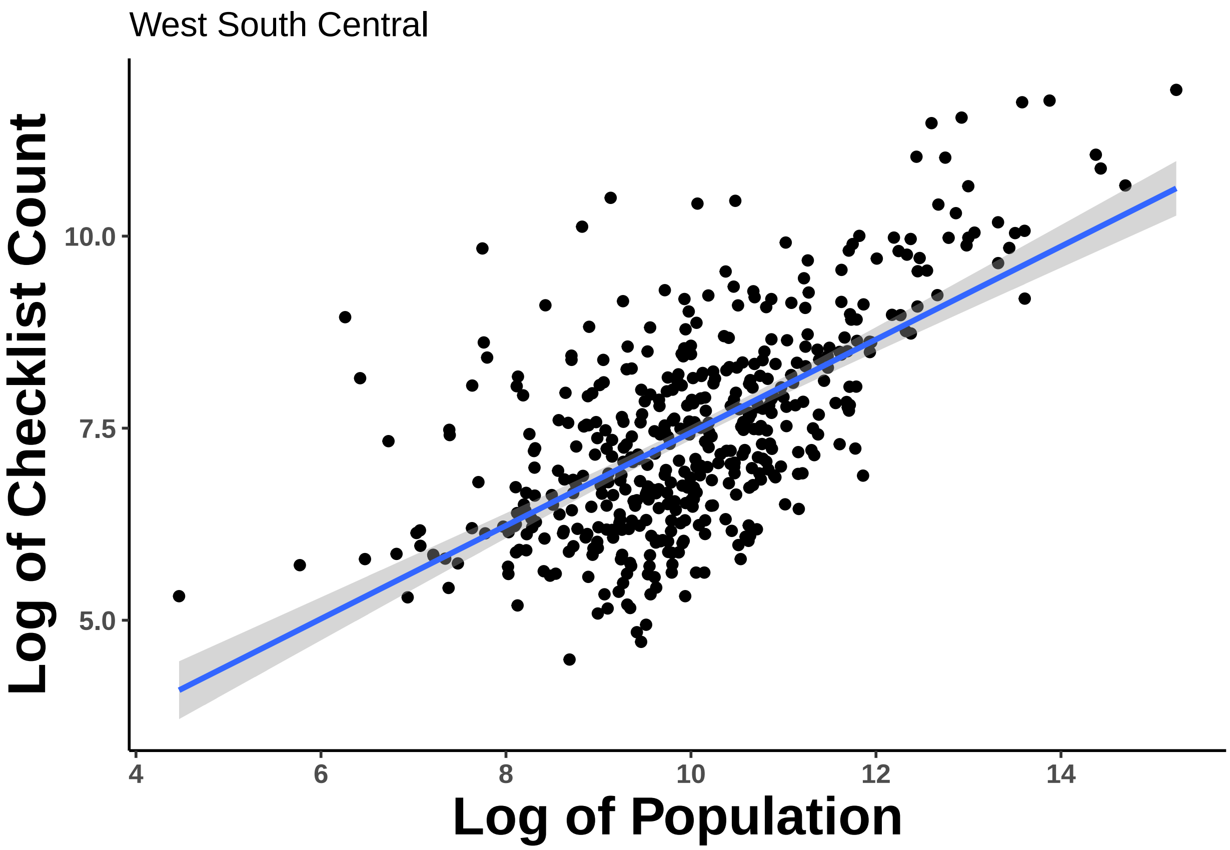

```
## [1] "Relationship undefined: y=(0.639x +/- 0.191) + (2.68 +/- 1.82), r^2 = 0.6"
## Saving 6.5 x 4.5 in image
```



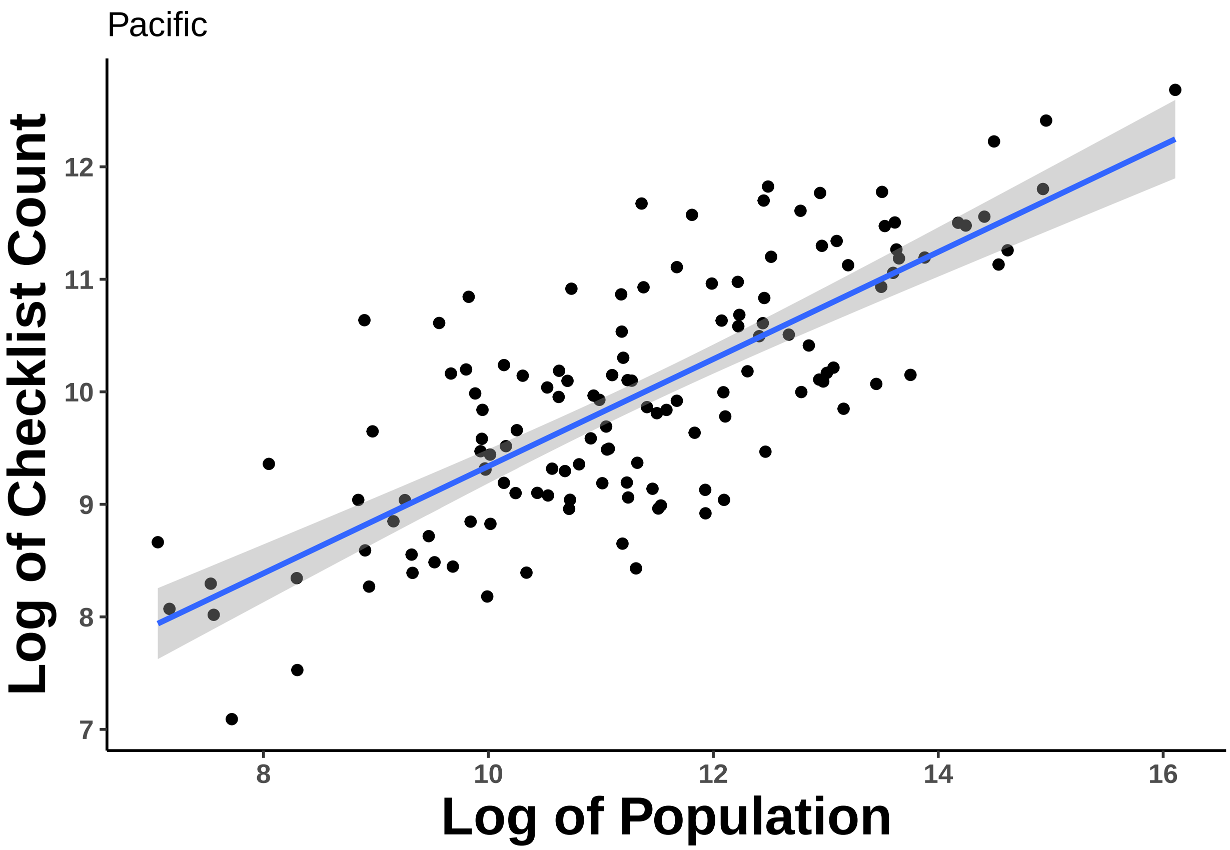

```
## [1] "Relationship Mountain: y=(0.635x +/- 0.062) + (2.22 +/- 0.62), r^2 = 0.6"
## Saving 6.5 x 4.5 in image
```




```
## [1] "Relationship West South Central: y=(0.606x +/- 0.067) + (1.38 +/- 0.68), r^2 = 0.41"
## Saving 6.5 x 4.5 in image
```

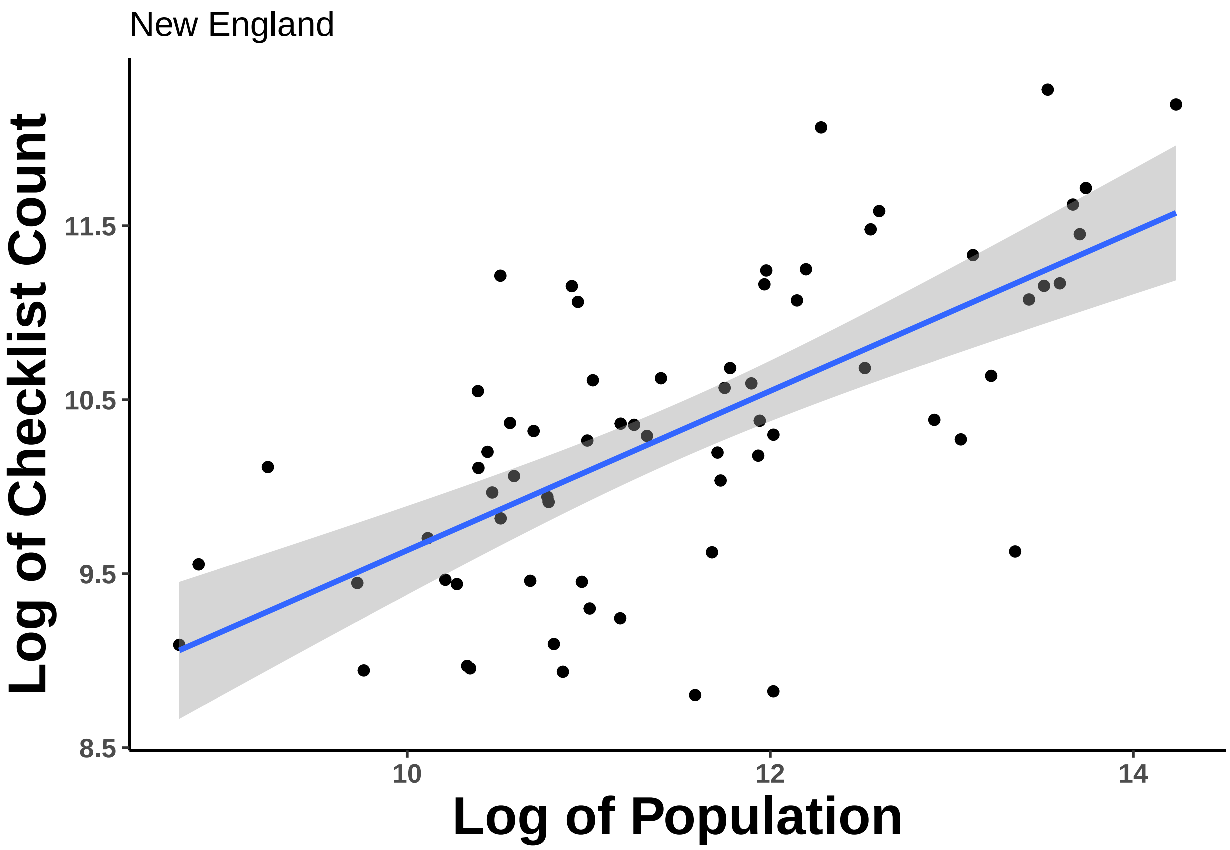



```
## [1] "Relationship Pacific: y=(0.476x +/- 0.069) + (4.58 +/- 0.8), r^2 = 0.59"

## Saving 6.5 x 4.5 in image
```

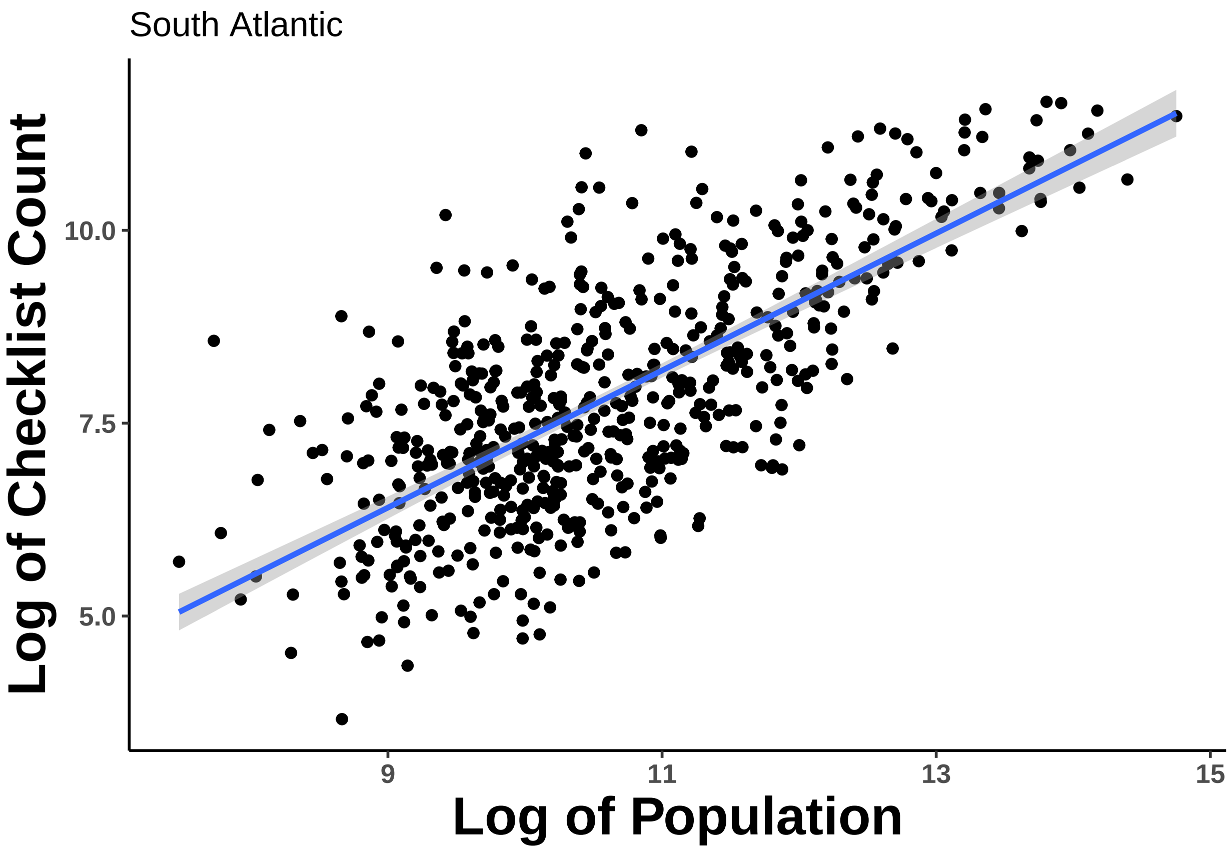



```
## [1] "Relationship New England: y=(0.458x +/- 0.13) + (5.06 +/- 1.5), r^2 = 0.43"
```

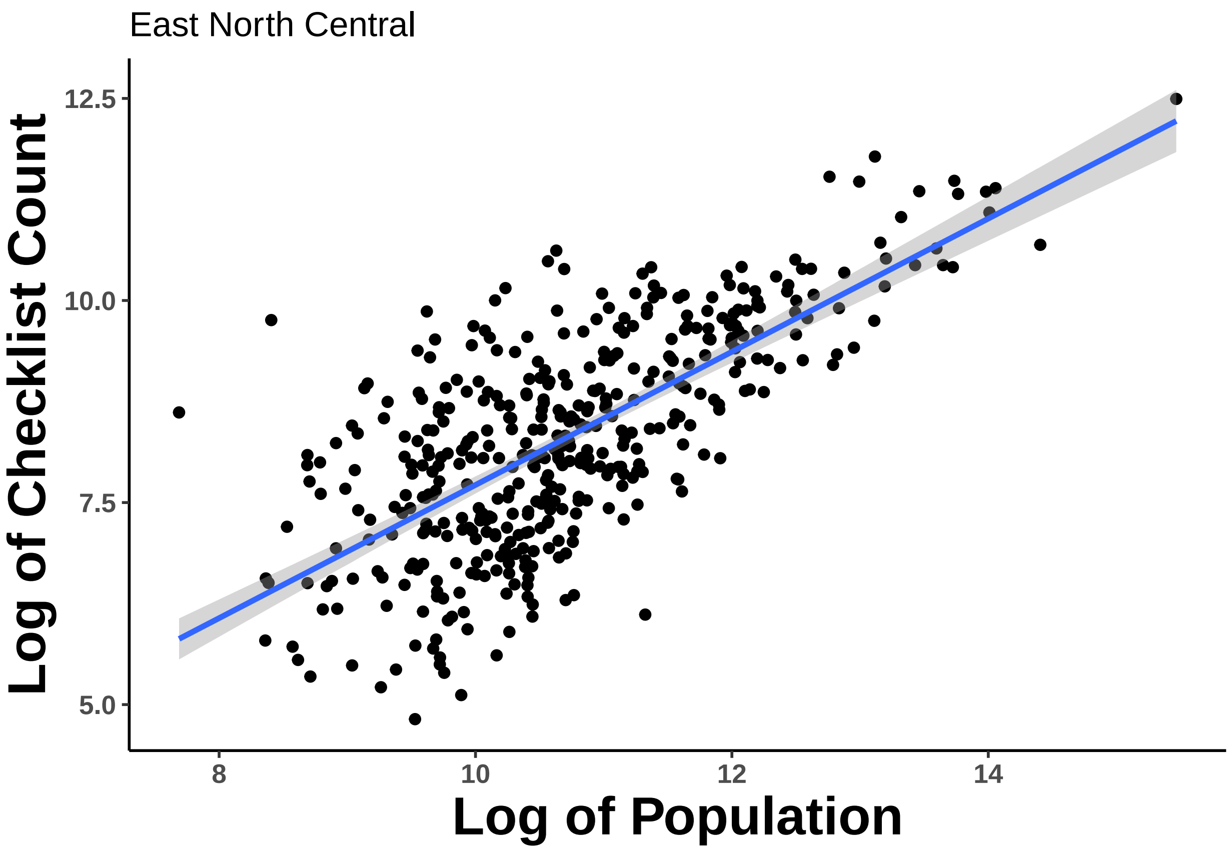
```
## Saving 6.5 x 4.5 in image
```



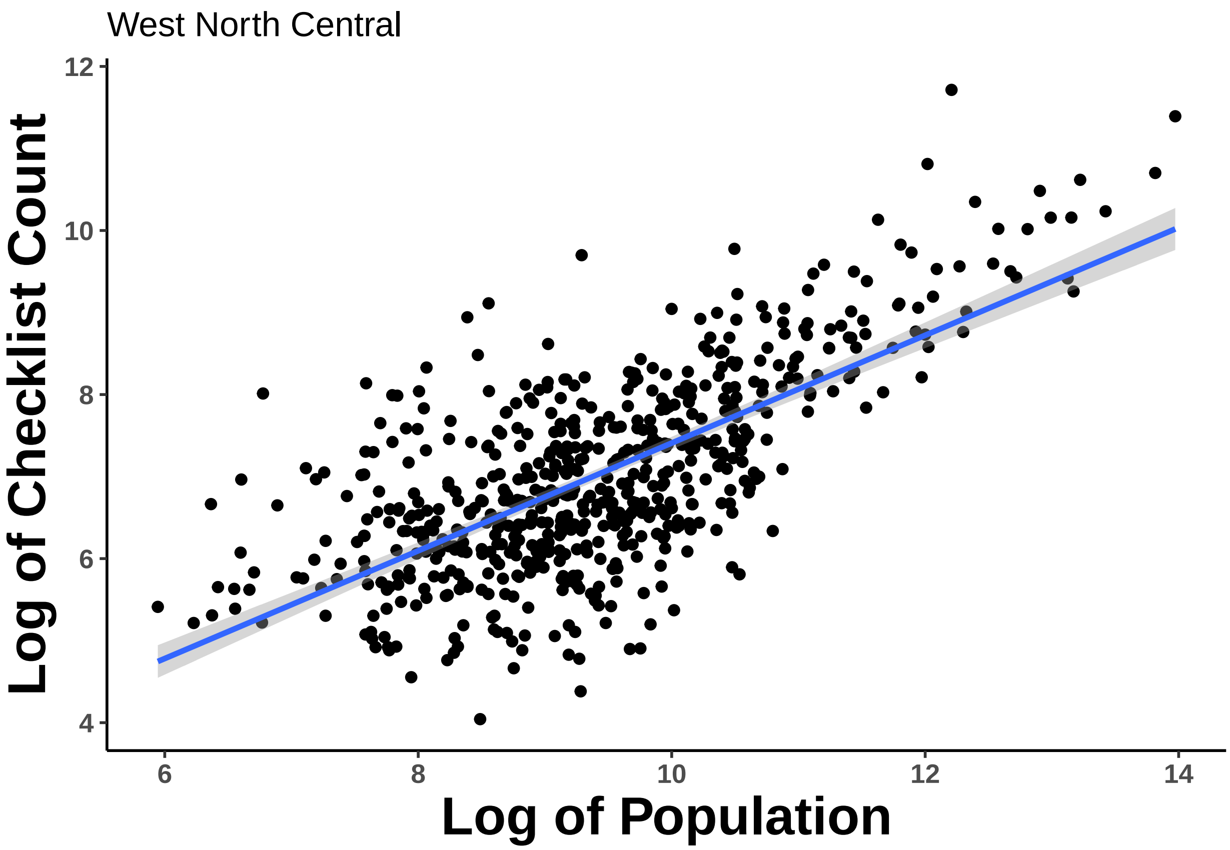

```
## [1] "Relationship South Atlantic: y=(0.889x +/- 0.071) + (-1.59 +/- 0.76), r^2 = 0.51"
## Saving 6.5 x 4.5 in image
```

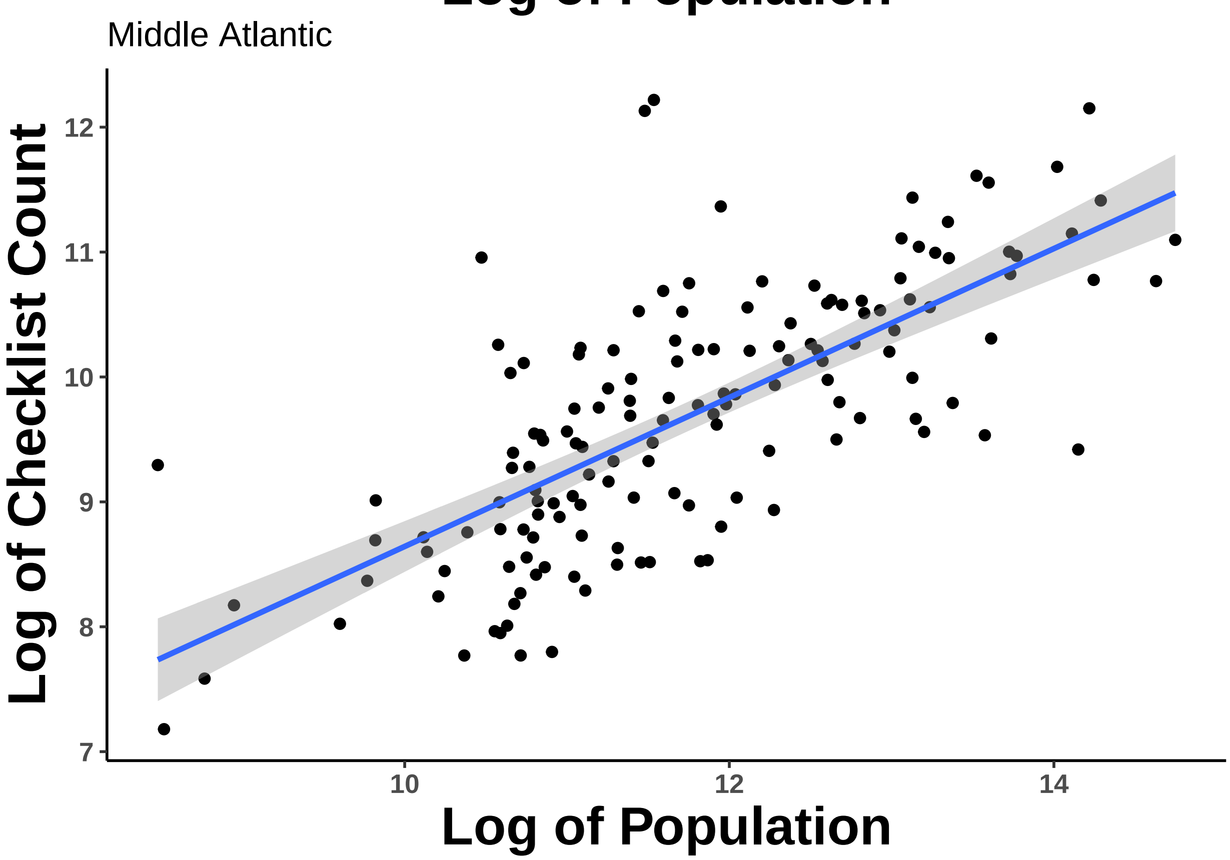



```
## [1] "Relationship East North Central: y=(0.824x +/- 0.08) + (-0.52 +/- 0.86), r^2 = 0.49"
## Saving 6.5 x 4.5 in image
```




```
## [1] "Relationship West North Central: y=(0.657x +/- 0.055) + (0.84 +/- 0.52), r^2 = 0.48"
## Saving 6.5 x 4.5 in image
```

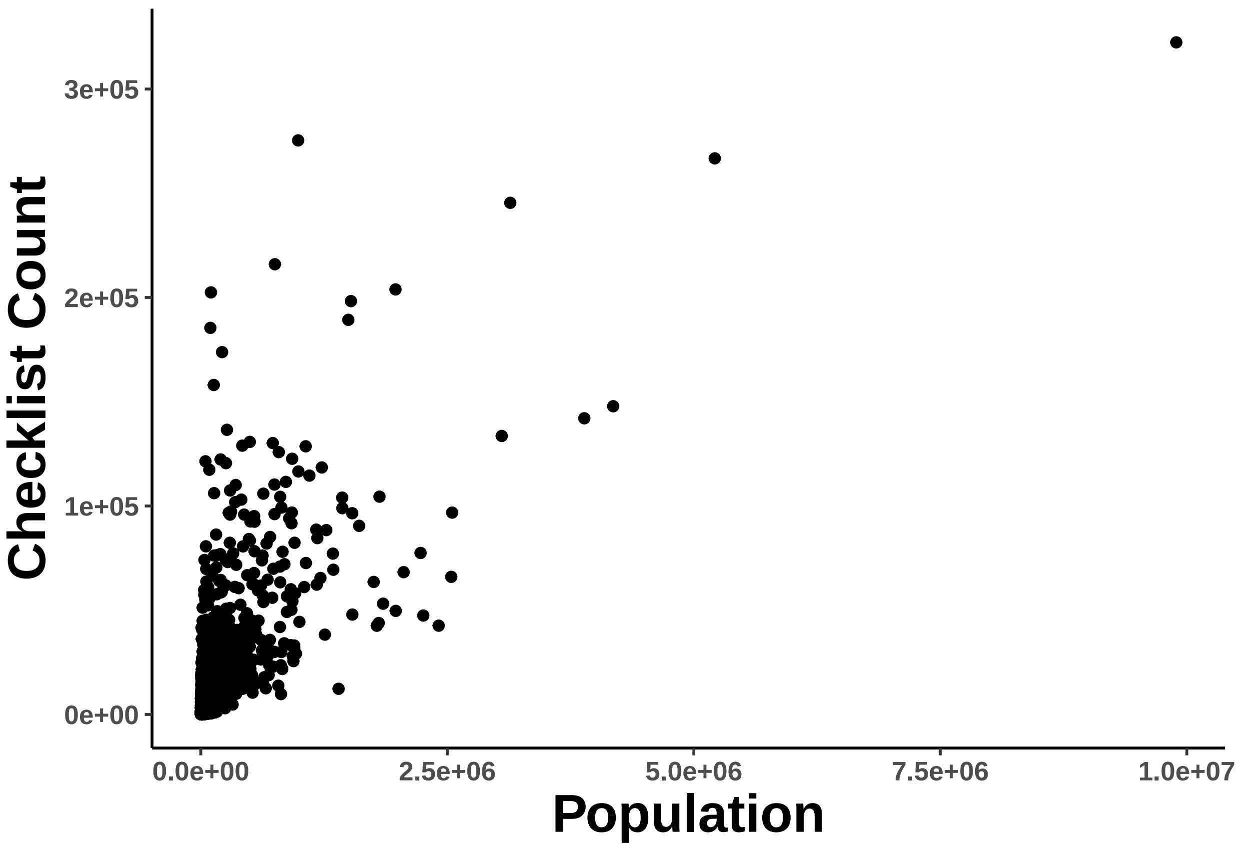





```
## [1] "Relationship Middle Atlantic: y=(0.596x +/- 0.096) + (2.68 +/- 1.14), r^2 = 0.51"
```

Overview of all data

```
a=ggplot(aes(x=Population,y=eBird_count),data=x)
b=geom_point()
#c=geom_smooth(method=lm)
d=theme_classic()
d.1=theme(axis.title = element_text(face="bold",size=20),
          axis.text = element_text(size=10,face="bold"),
          legend.title = element_blank(),
          legend.text = element_text(size=14,face="bold"))
#w2=ggtitle(subregions[i])

e=labs(x='Population',y='Checklist Count')

Fig3=a+b+d+e+d.1

plot(Fig3)
```




```
#ggsave(Fig3,paste0(filepath,"Non-transformed_pop-checklist.png"),dpi=400)
```