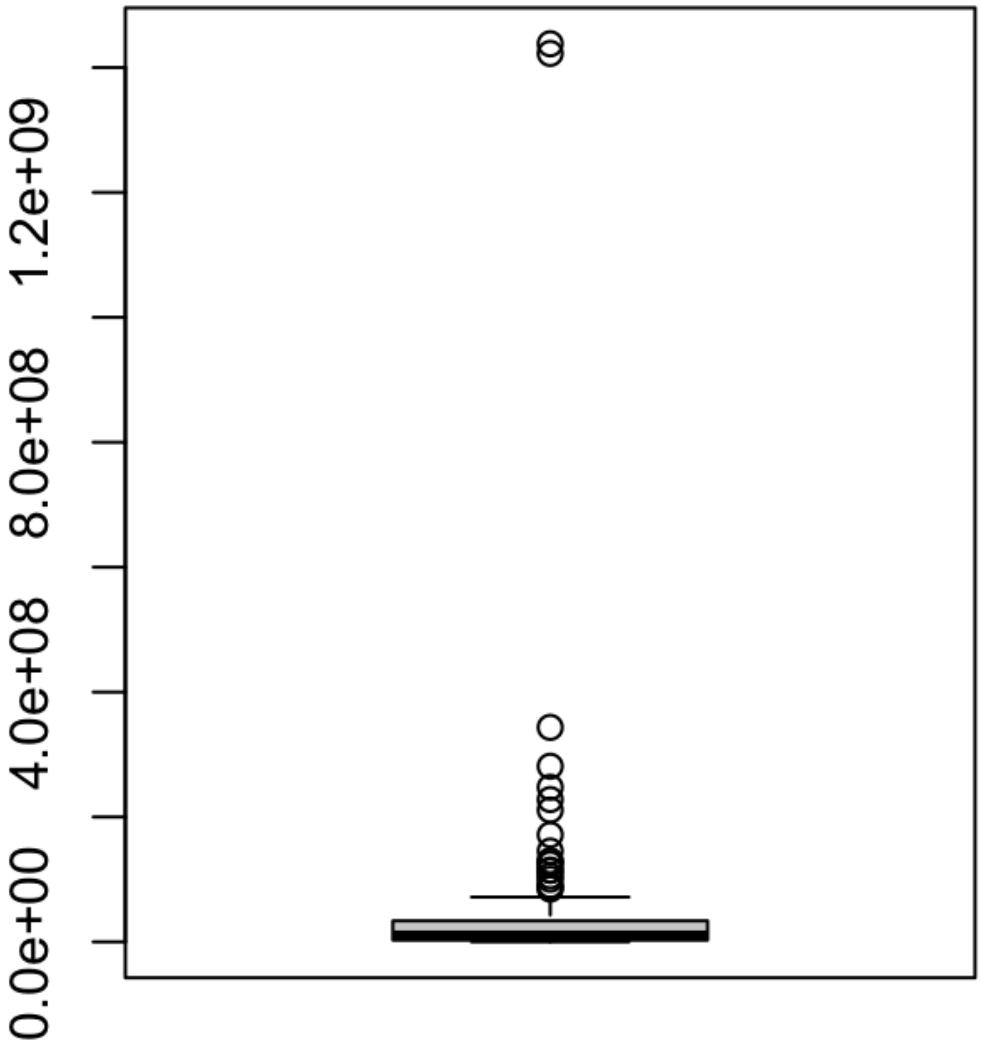
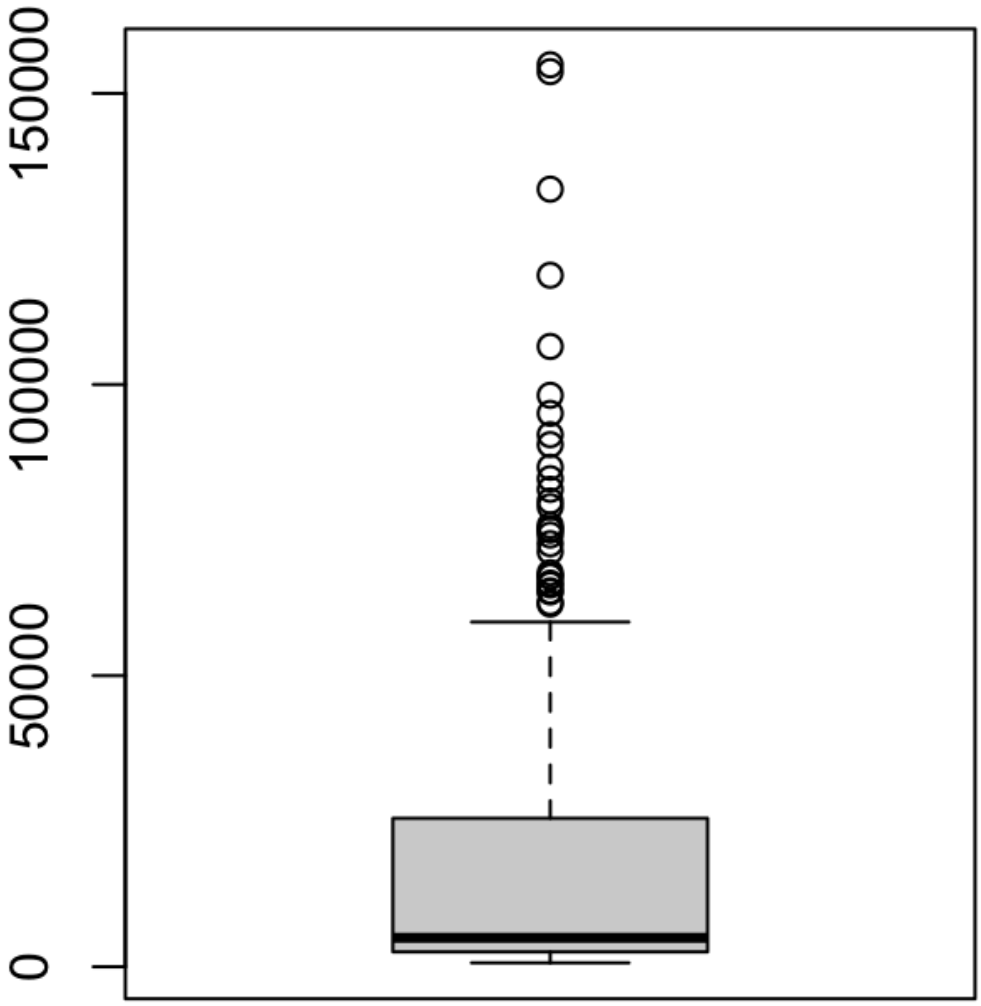
Sean McHugh
September 12, 2025
Lab 1

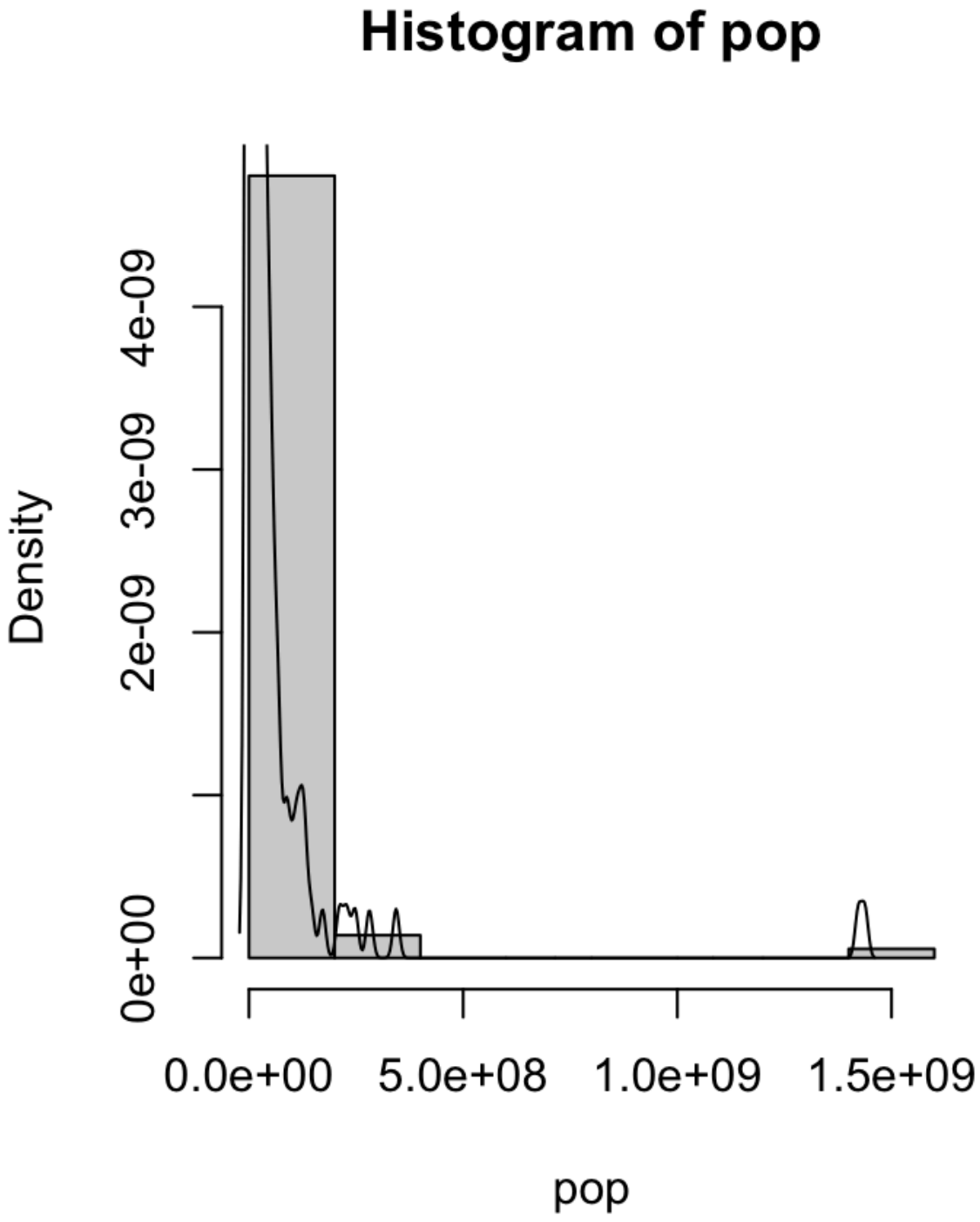
```
 ## Choose variables "population" and "gdp"
> pop <- epi.data$population
> gdp <- epi.data$gdp
>
> # Variable summaries
> summary(pop)
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.      NA's
3.883e+04 2.755e+06 1.024e+07 4.436e+07 3.371e+07 1.438e+09         1
> summary(gdp)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
    671    2552    4970   21470   25505  154915       1
>
> # Variable boxplots
> boxplot(pop, name = "Population")
```




```
> boxplot(gdp, name = "GDP")
```



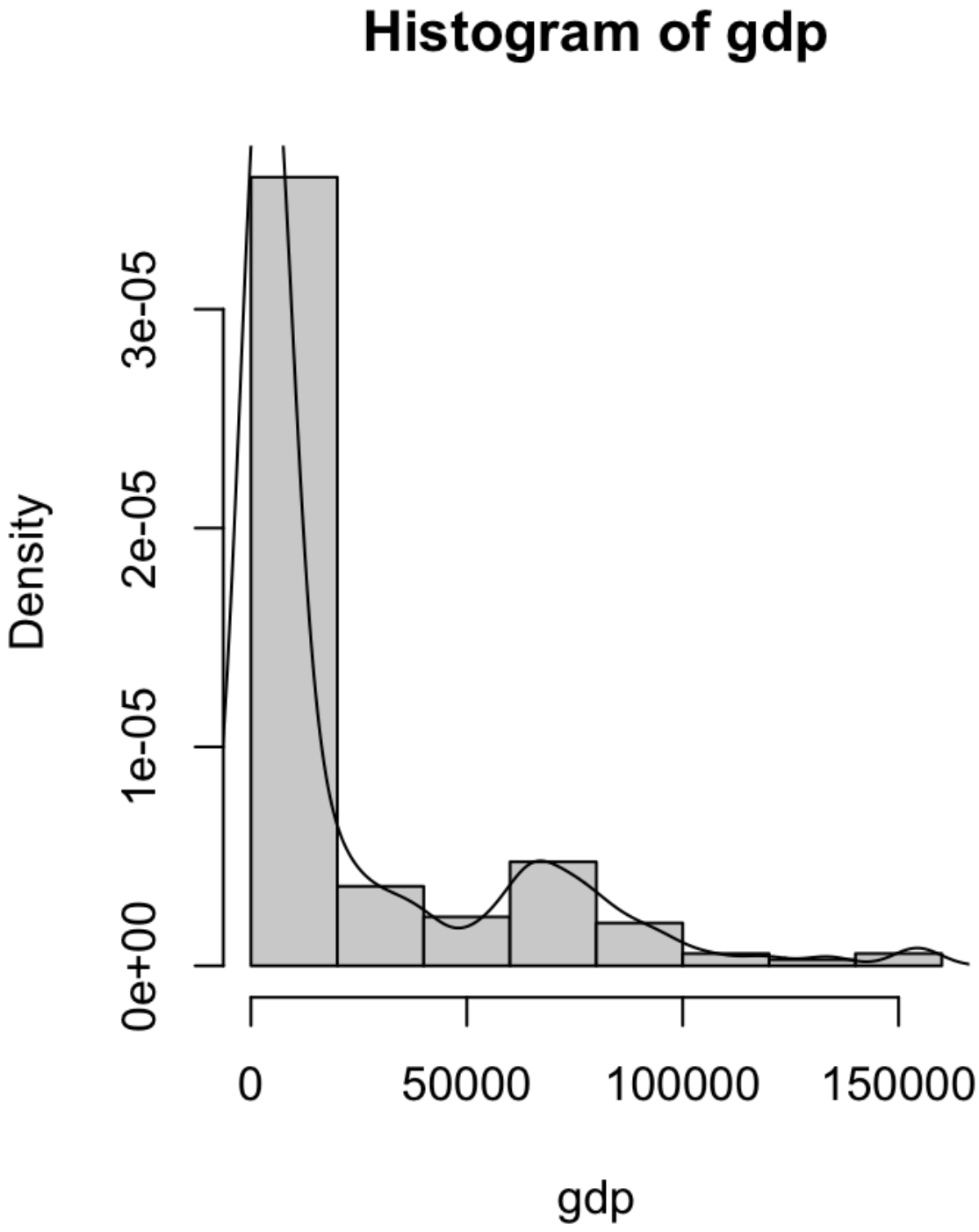
>
> # Histograms over theoretical probability distributions
> hist(pop, prob=TRUE)
> lines(density(pop,na.rm=TRUE))



>

> hist(gdp, prob=TRUE)

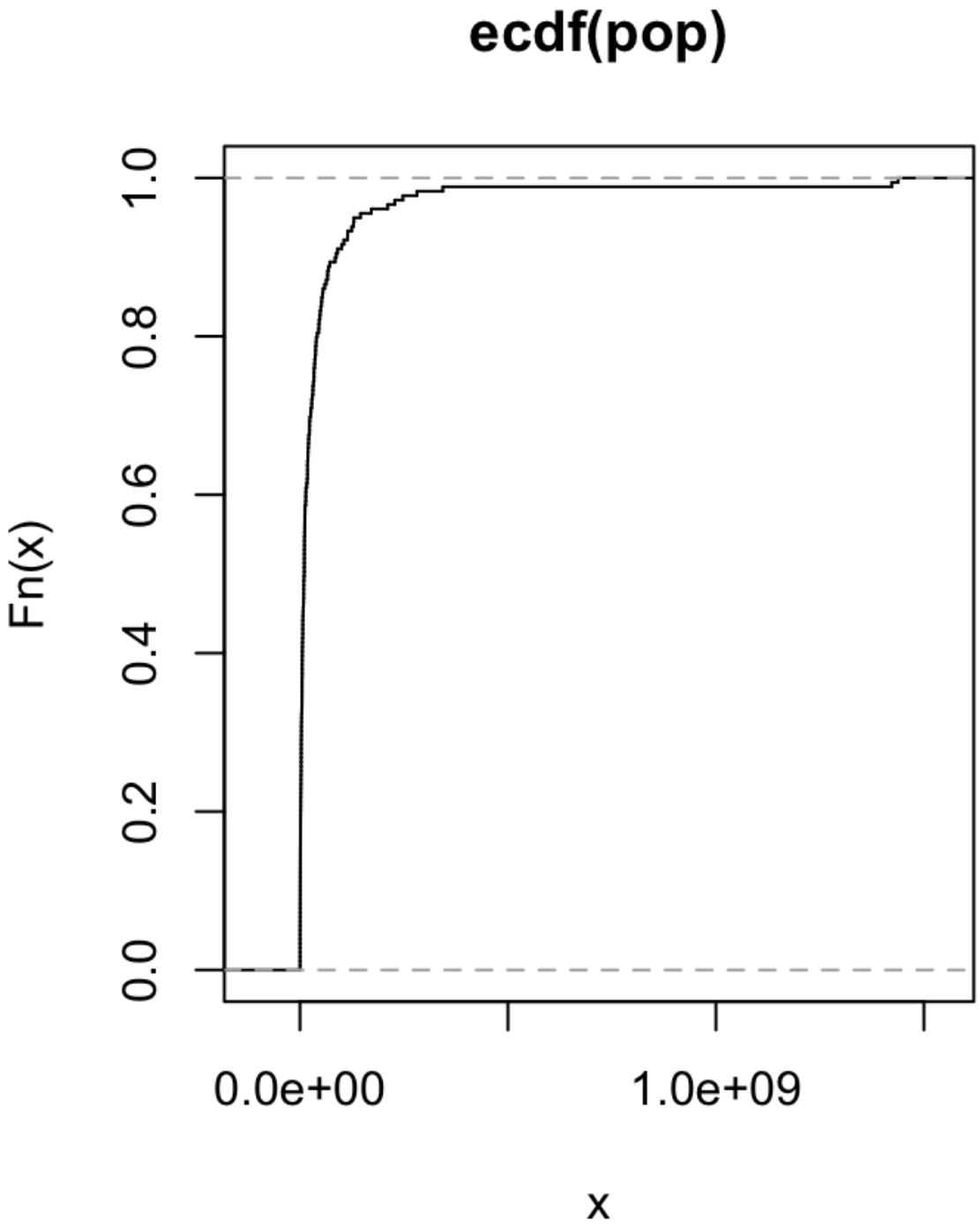
> lines(density(gdp,na.rm=TRUE))



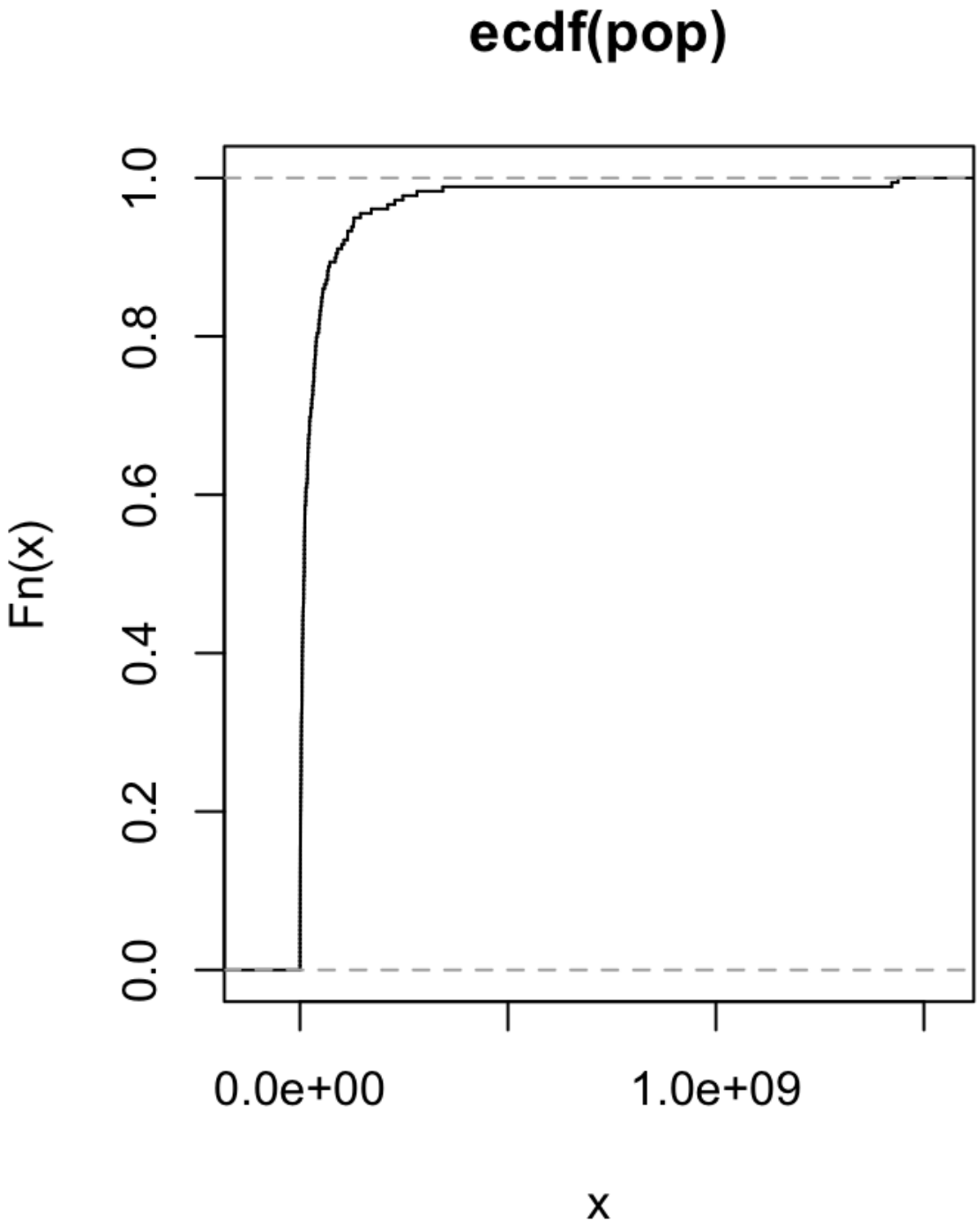
>

> # QQ plots of each variable against normal distribution

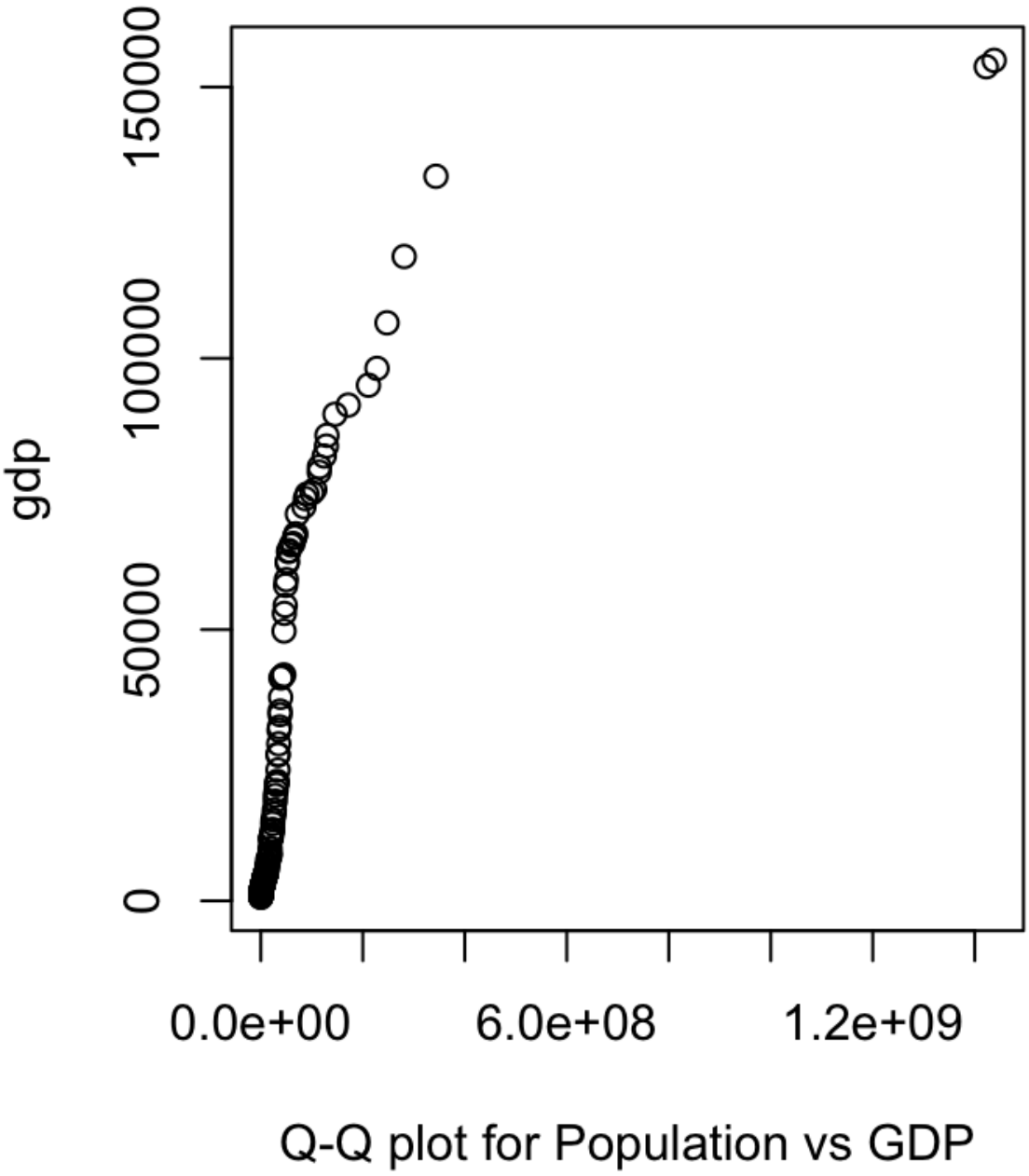
> plot(ecdf(pop), do.points=FALSE, verticals=TRUE)



> plot(ecdf(gdp), do.points=FALSE, verticals=TRUE)



>
> # QQ plot of the 2 variables against each other
> qqplot(pop, gdp, xlab = "Q-Q plot for Population vs GDP")



>
> ## Normality statsitical test for each variable
> shapiro.test(pop)

```
        Shapiro-Wilk normality test

data:  pop
W = 0.24368, p-value < 2.2e-16

> shapiro.test(gdp)

        Shapiro-Wilk normality test

data:  gdp
W = 0.66982, p-value < 2.2e-16

>
> ## Install package for ad test
> install.packages("nortest")
trying URL 'https://cran.rstudio.com/bin/macosx/big-sur-arm64/contrib/4.5/nortest_1.0-4.tgz'
Content type 'application/x-gzip' length 37471 bytes (36 KB)
==================================================
downloaded 36 KB


The downloaded binary packages are in
        /var/folders/hz/qx4103vj5y9gjdcfyy05y8fm0000gn/T//Rtmpiqyudk/downloaded_packages
> library(nortest)
>
> ad.test(pop)

        Anderson-Darling normality test

data:  pop
A = 42.815, p-value < 2.2e-16

> ad.test(gdp)

        Anderson-Darling normality test

data:  gdp
A = 24.234, p-value < 2.2e-16

>
> # Statistical test for variables having identical distributions
> ks.test(pop,gdp)

        Asymptotic two-sample Kolmogorov-Smirnov test
```

```
data:  pop and gdp
D = 0.95531, p-value < 2.2e-16
alternative hypothesis: two-sided

>
> wilcox.test(pop,gdp)

        Wilcoxon rank sum test with continuity correction

data:  pop and gdp
W = 31950, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0

>
> var.test(pop,gdp)

        F test to compare two variances

data:  pop and gdp
F = 23637964, num df = 178, denom df = 178, p-value < 2.2e-16
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 17603539 31740968
sample estimates:
ratio of variances
         23637964

> t.test(pop,gdp)

        Welch Two Sample t-test

data:  pop and gdp
t = 3.7956, df = 178, p-value = 0.0002017
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 21286903 67391845
sample estimates:
 mean of x  mean of y
44360844.0    21470.1
```