```
> ####### Data Analytics Fall 2024 Lab 02 ######

> library(ggplot2)

> ### set working directory

> setwd("~/Courses/Data Analytics/Fall24/labs/lab01/")

Error in setwd("~/Courses/Data Analytics/Fall24/labs/lab01/") :

  cannot change working directory

> ### read in data

> epi.results <- read.csv("C:/Users/chaos/Downloads/epi2024results06022024.csv", header=TRUE)

> epi.weights <- read.csv("C:/Users/chaos/Downloads/epi2024weights.csv")

> View(epi.results)

> View(epi.weights)

> epi.results$EPI.new

 [1] 30.7 52.1 41.9 39.7 55.5 46.8 44.7 63.0 69.0 40.4 56.0 35.9 27.8 53.1 58.1 66.7

 [17] 47.4 37.4 43.3 44.9 45.6 49.0 53.0 48.5 56.3 41.5 33.0 37.9 31.0 38.1 61.1 38.3

 [33] 35.2 50.0 35.5 49.4 37.9 55.5 42.5 62.6 52.3 54.0 65.6 39.0 67.9 32.2 49.2 47.6

 [49] 51.2 43.8 41.5 41.6 28.6 75.3 38.5 35.8 45.8 73.7 67.1 53.1 37.1 46.9 74.6 36.6

 [65] 67.4 46.0 32.6 36.2 41.6 48.6 36.2 40.2 60.1 64.3 27.6 33.8 41.6 30.4 65.7 48.1

 [81] 60.5 48.5 61.7 47.5 47.5 36.9 44.1 44.9 42.2 26.1 59.9 40.1 36.6 34.1 63.9 75.0

 [97] 29.9 34.9 41.2 38.1 33.9 66.6 42.6 34.2 47.3 44.7 40.6 45.6 37.0 47.6 39.7 38.6

[113] 26.9 43.8 32.9 67.2 57.7 47.4 39.2 37.5 50.0 70.0 51.9 25.5 52.9 36.5 39.0 46.6

[129] 32.0 64.4 62.2 47.2 41.2 57.2 46.5 33.4 51.0 54.1 46.8 35.9 42.6 43.3 49.3 48.2

[145] 39.7 53.8 65.0 62.5 41.8 42.9 51.0 64.2 38.7 38.6 56.6 70.5 68.0 50.3 31.9 43.1

[161] 45.4 49.7 35.2 40.2 52.1 45.7 37.6 40.7 35.4 54.6 52.0 72.7 57.3 43.9 42.9 44.6

[177] 53.1 24.5 46.1 51.7

> epi.results[1,5]
```

```
[1] 30.7

> attach(epi.results)

> EPI.new

 [1] 30.7 52.1 41.9 39.7 55.5 46.8 44.7 63.0 69.0 40.4 56.0 35.9 27.8 53.1 58.1 66.7

[17] 47.4 37.4 43.3 44.9 45.6 49.0 53.0 48.5 56.3 41.5 33.0 37.9 31.0 38.1 61.1 38.3

[33] 35.2 50.0 35.5 49.4 37.9 55.5 42.5 62.6 52.3 54.0 65.6 39.0 67.9 32.2 49.2 47.6

[49] 51.2 43.8 41.5 41.6 28.6 75.3 38.5 35.8 45.8 73.7 67.1 53.1 37.1 46.9 74.6 36.6

[65] 67.4 46.0 32.6 36.2 41.6 48.6 36.2 40.2 60.1 64.3 27.6 33.8 41.6 30.4 65.7 48.1

[81] 60.5 48.5 61.7 47.5 47.5 36.9 44.1 44.9 42.2 26.1 59.9 40.1 36.6 34.1 63.9 75.0

[97] 29.9 34.9 41.2 38.1 33.9 66.6 42.6 34.2 47.3 44.7 40.6 45.6 37.0 47.6 39.7 38.6

[113] 26.9 43.8 32.9 67.2 57.7 47.4 39.2 37.5 50.0 70.0 51.9 25.5 52.9 36.5 39.0 46.6

[129] 32.0 64.4 62.2 47.2 41.2 57.2 46.5 33.4 51.0 54.1 46.8 35.9 42.6 43.3 49.3 48.2

[145] 39.7 53.8 65.0 62.5 41.8 42.9 51.0 64.2 38.7 38.6 56.6 70.5 68.0 50.3 31.9 43.1

[161] 45.4 49.7 35.2 40.2 52.1 45.7 37.6 40.7 35.4 54.6 52.0 72.7 57.3 43.9 42.9 44.6

[177] 53.1 24.5 46.1 51.7

> EPI.new[1]

[1] 30.7

> ## NA values

> na.indices <- is.na(EPI.new)

> ## drop NAs

> Epi.new.compl <- EPI.new[!na.indices]

> ## convert to data frame and add country

> Epi.new.compl <- data.frame(Country = country[!na.indices], EPI = EPI.new[!na.indices])

> ## summary stats

> summary(EPI.new)
```

```
   Min. 1st Qu.  Median   Mean 3rd Qu.   Max.
  24.50   38.25   45.50   46.84   53.10   75.30
> fivenum(EPI.new,na.rm=TRUE)
[1] 24.5 38.2 45.5 53.1 75.3
> ## histograms
> hist(EPI.new)
> hist(EPI.new, seq(20., 80., 2.0), prob=TRUE)
> rug(EPI.new)
> lines(density(EPI.new,na.rm=TRUE,bw=1))
> lines(density(EPI.new,na.rm=TRUE,bw="SJ"))
> x <- seq(20., 80., 1.0)
> qn<- dnorm(x,mean=42, sd=5,log=FALSE)
> lines(x,0.4*qn)
> qn<- dnorm(x,mean=65, sd=5,log=FALSE)
> lines(x,0.12*qn)
> boxplot(EPI.old, EPI.new, names=c("EPI.old","EPI.new"))
> qqnorm(EPI.new)
> qqline(EPI.new)
> x <- seq(20., 80., 1.0)
> qqplot(qnorm(ppoints(200)), x)
> qqline(x)
> qqplot(qnorm(ppoints(200)),EPI.new)
> qqline(EPI.new)
> qqplot(rnorm(1000),EPI.new)
> qqline(EPI.new)
```

```
> d1 <- rnorm(10000)

> d2 <- rnorm(10000)

> qqplot(d1,d1)

> qqline(d1)

> ### Empirical Cumulative Distribution Function

> plot(ecdf(EPI.new), do.points=FALSE)

> plot(ecdf(rnorm(1000, 45, 10)), do.points=FALSE, main="Norm Dist vs. EPI.new ECDF")

> lines(ecdf(EPI.new))

> plot(ecdf(EPI.old), do.points=FALSE, main="EPI.old vs. EPI.new ECDF")

> lines(ecdf(EPI.new))

> ## read data

> populations_2023 <- read.csv("C:/Users/chaos/Downloads/countries_populations_2023.csv")

> ## drop country populations that don't exist in epi results

> populations <- populations_2023[-which(!populations_2023$Country %in% epi.results$country),]

> ## sort populations by country name

> populations <- populations[order(populations$Country),]

> ## drop country results that don't exist in populations

> epi.results.sub <- epi.results[-which(!epi.results$country %in% populations$Country),]

> ## sort results by country name

> epi.results.sub <- epi.results.sub[order(epi.results.sub$country),]

> ## only keep relevant columns

> epi.results.sub <- epi.results.sub[,c("country","EPI.old","EPI.new")]

> ## convert to mnumeric

> epi.results.sub$population <- as.numeric(populations$Population)

> ## compute population log
```

```
> epi.results.sub$population_log <- log10(epi.results.sub$population)

> boxplot(epi.results.sub$population_log)

> attach(epi.results.sub)

The following objects are masked from epi.results:


    country, EPI.new, EPI.old


> ## created linear model of EPI.new = a(population_log) + b

> lin.mod.epinew <- lm(EPI.new~population_log,epi.results.sub)

> plot(EPI.new~population_log)

> abline(lin.mod.epinew)

> summary(lin.mod.epinew)


Call:

lm(formula = EPI.new ~ population_log, data = epi.results.sub)


Residuals:

   Min    1Q Median    3Q    Max

-21.017 -8.608 -2.396  6.046 29.789


Coefficients:

        Estimate Std. Error t value Pr(>|t|)

(Intercept)    62.340    7.587  8.216 6.17e-14 ***

population_log  -2.211    1.087 -2.035  0.0435 *

---
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11.58 on 163 degrees of freedom

Multiple R-squared: 0.02478,   Adjusted R-squared: 0.01879

F-statistic: 4.141 on 1 and 163 DF, p-value: 0.04348


```
> plot(lin.mod.epinew)

Hit <Return> to see next plot: ggplot(epi.results.sub, aes(x = population_log, y = EPI.new)) +

Hit <Return> to see next plot:   geom_point() +

Hit <Return> to see next plot:   stat_smooth(method = "lm")

Hit <Return> to see next plot: ggplot(lin.mod.epinew, aes(x = .fitted, y = .resid)) +

>  geom_point() +

+  geom_hline(yintercept = 0) +

+  labs(title='Residual vs. Fitted Values Plot', x='Fitted Values', y='Residuals')
```

Error in `+.gg`:

! Cannot add <ggproto> objects together.

i Did you forget to add this object to a <ggplot> object?

Run `rlang::last_trace()` to see where the error occurred.

```
> ggplot(lin.mod.epinew, aes(x = .fitted, y = .resid)) +

+  geom_point() +

+  geom_hline(yintercept = 0) +

+  labs(title='Residual vs. Fitted Values Plot', x='Fitted Values', y='Residuals')

> ## another lm

> lin.mod.pop <- lm(population_log~EPI.new,epi.results.sub)

> plot(population_log~EPI.old)
```

```
> abline(lin.mod.pop)

> summary(lin.mod.pop)


Call:

lm(formula = population_log ~ EPI.new, data = epi.results.sub)


Residuals:

    Min     1Q  Median     3Q    Max

-2.39347 -0.46967  0.07525  0.53462  2.09092


Coefficients:

          Estimate Std. Error t value Pr(>|t|)

(Intercept)  7.459928   0.266680  27.973  <2e-16 ***

EPI.new     -0.011205   0.005506  -2.035  0.0435 *

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Residual standard error: 0.8246 on 163 degrees of freedom

Multiple R-squared:  0.02478,      Adjusted R-squared:  0.01879

F-statistic: 4.141 on 1 and 163 DF,  p-value: 0.04348


> plot(lin.mod.pop)

Hit <Return> to see next plot: ggplot(epi.results.sub, aes(x = EPI.old, y = population_log)) +

Hit <Return> to see next plot:   geom_point() +

Hit <Return> to see next plot:   stat_smooth(method = "lm")
```

Hit <Return> to see next plot: ggplot(lin.mod.pop, aes(x = .fitted, y = .resid)) +

> geom_point() +

+ geom_hline(yintercept = 0) +

+ labs(title='Residual vs. Fitted Values Plot', x='Fitted Values', y='Residuals')

Error in `+.gg` :

! Cannot add <ggproto> objects together.

i Did you forget to add this object to a <ggplot> object?

Run `rlang::last_trace()` to see where the error occurred.

> ggplot(lin.mod.pop, aes(x = .fitted, y = .resid)) +

+ geom_point() +

+ geom_hline(yintercept = 0) +

+ labs(title='Residual vs. Fitted Values Plot', x='Fitted Values', y='Residuals')

> #Exercise 1: fitting a distribution

> #Variable 1: SHI

> epi.results$SHI.new

```
  [1] 66.4 53.5 74.0 71.6   NA 48.2 80.4 50.3 59.9 80.1 98.2   NA 80.1
 [14]   NA 77.8 48.0  0.0 80.4 87.7 24.3 75.4 94.2  0.0 33.3 77.5 74.9
 [27] 80.7   NA  0.0  0.0 47.1 77.5 70.2 46.8 24.9  0.0 30.4 14.0  0.0
 [40] 78.4 21.6 60.2 68.1  0.0 87.1 81.3   NA 56.7 30.1 90.1 66.7  8.8
 [53] 86.3 43.3 26.0 78.1 64.9 13.5 72.8 77.8 41.8 85.7 71.9 42.1 85.1
 [66]   NA  0.0 25.4  0.0 79.8 22.2  0.0 77.2 100.0 61.1  0.0 69.6 87.1
 [79] 97.7  0.0 59.4  0.0 43.3 62.3 88.9 80.7   NA 67.5 83.9  0.0 44.2
 [92] 63.2 53.2  0.0 77.5 66.4  0.0 86.5  0.0   NA 85.4   NA   NA 80.4
[105]   NA  8.2   NA 58.8 58.5 53.2 49.7 67.0 17.0 77.2 95.9 52.9 59.4
[118]  0.0 72.2 62.0 67.0 74.0 61.1 32.7  0.0 50.0  0.0 10.5  0.0 83.6
```

```
[131] 64.0 66.1 72.8 62.3 67.3 67.3   NA   NA 100.0   NA 89.2 89.2 84.5

[144]   NA  0.0   NA 73.7 52.6  0.0 77.2 35.4 67.3 30.7 70.2 74.9  3.5

[157] 68.7 27.8 87.7 73.4  0.0 69.3 81.0   NA 20.8 71.9 64.0 89.2 63.7

[170] 71.1 78.7 88.0 20.8 69.9 66.1 98.8 56.7  0.0 80.1 85.1

> epi.results[1,5]

[1] 30.7

> attach(epi.results)
```

The following objects are masked from epi.results.sub:

   country, EPI.new, EPI.old

The following objects are masked from epi.results (pos = 4):

  AGR.new, AGR.old, AIR.new, AIR.old, APO.new, APO.old, BCA.new, BCA.old,

  BDH.new, BDH.old, BER.new, BER.old, BTO.new, BTO.old, BTZ.new, BTZ.old,

  CBP.new, CBP.old, CCH.new, CCH.old, CDA.new, CDA.old, CDF.new, CDF.old,

  CHA.new, CHA.old, code, COE.new, COE.old, country, ECO.new, ECO.old,

  ECS.new, ECS.old, EPI.new, EPI.old, FCD.new, FCD.old, FCL.new, FCL.old,

  FGA.new, FGA.old, FLI.new, FLI.old, FSH.new, FSH.old, FSS.new, FSS.old,

  GHN.new, GHN.old, GTI.new, GTI.old, GTP.new, GTP.old, H2O.new, H2O.old,

  HFD.new, HFD.old, HLT.new, HLT.old, HMT.new, HMT.old, HPE.new, HPE.old,

  IFL.new, IFL.old, iso, LED.new, LED.old, LUF.new, LUF.old, MHP.new,

  MHP.old, MKP.new, MKP.old, MPE.new, MPE.old, NDA.new, NDA.old, NOD.new,

  NOD.old, NXA.new, NXA.old, OEB.new, OEB.old, OEC.new, OEC.old, OZD.new,

  OZD.old, PAE.new, PAE.old, PAR.new, PAR.old, PCC.new, PCC.old, PFL.new,

PFL.old, PHL.new, PHL.old, PRS.new, PRS.old, PSU.new, PSU.old, RCY.new,

RCY.old, RLI.new, RLI.old, RMS.new, RMS.old, SDA.new, SDA.old, SHI.new,

SHI.old, SMW.new, SMW.old, SNM.new, SNM.old, SOE.new, SOE.old, SPI.new,

SPI.old, TBN.new, TBN.old, TCG.new, TCG.old, TKP.new, TKP.old, USD.new,

USD.old, UWD.new, UWD.old, VOE.new, VOE.old, WMG.new, WMG.old, WPC.new,

WPC.old, WRR.new, WRR.old, WRS.new, WRS.old, WWC.new, WWC.old, WWG.new,

WWG.old, WWR.new, WWR.old, WWT.new, WWT.old

```
> SHI.new
  [1] 66.4 53.5 74.0 71.6   NA 48.2 80.4 50.3 59.9 80.1 98.2   NA 80.1
 [14]   NA 77.8 48.0  0.0 80.4 87.7 24.3 75.4 94.2  0.0 33.3 77.5 74.9
 [27] 80.7   NA  0.0  0.0 47.1 77.5 70.2 46.8 24.9  0.0 30.4 14.0  0.0
 [40] 78.4 21.6 60.2 68.1  0.0 87.1 81.3   NA 56.7 30.1 90.1 66.7  8.8
 [53] 86.3 43.3 26.0 78.1 64.9 13.5 72.8 77.8 41.8 85.7 71.9 42.1 85.1
 [66]   NA  0.0 25.4  0.0 79.8 22.2  0.0 77.2 100.0 61.1  0.0 69.6 87.1
 [79] 97.7  0.0 59.4  0.0 43.3 62.3 88.9 80.7   NA 67.5 83.9  0.0 44.2
 [92] 63.2 53.2  0.0 77.5 66.4  0.0 86.5  0.0   NA 85.4   NA   NA 80.4
[105]   NA  8.2   NA 58.8 58.5 53.2 49.7 67.0 17.0 77.2 95.9 52.9 59.4
[118]  0.0 72.2 62.0 67.0 74.0 61.1 32.7  0.0 50.0  0.0 10.5  0.0 83.6
[131] 64.0 66.1 72.8 62.3 67.3 67.3   NA   NA 100.0   NA 89.2 89.2 84.5
[144]   NA  0.0   NA 73.7 52.6  0.0 77.2 35.4 67.3 30.7 70.2 74.9  3.5
[157] 68.7 27.8 87.7 73.4  0.0 69.3 81.0   NA 20.8 71.9 64.0 89.2 63.7
[170] 71.1 78.7 88.0 20.8 69.9 66.1 98.8 56.7  0.0 80.1 85.1
> SHI.new[1]
[1] 66.4
```

```
> ## NA values

> na.indices <- is.na(SHI.new)

> ## drop NAs

> Shi.new.compl <- SHI.new[!na.indices]

> ## convert to data frame and add country

> Shi.new.compl <- data.frame(Country = country[!na.indices], SHI = SHI.new[!na.indices])

> ## summary stats

> summary(SHI.new)

  Min. 1st Qu.  Median   Mean 3rd Qu.   Max.   NA's

  0.00  28.38  64.45  53.83  77.80 100.00    18

> fivenum(SHI.new,na.rm=TRUE)

[1]  0.00  27.80  64.45  77.80 100.00

> ## histograms

> hist(SHI.new)

> hist(SHI.new, seq(0., 100., 2.0), prob=TRUE)

> rug(SHI.new)

> lines(density(SHI.new,na.rm=TRUE,bw=1))

> lines(density(SHI.new,na.rm=TRUE,bw="SJ"))

> x <- seq(0., 100., 1.0)

> qn<- dnorm(x,mean=42, sd=5,log=FALSE)

> lines(x,0.4*qn)

> qn<- dnorm(x,mean=65, sd=5,log=FALSE)

> lines(x,0.12*qn)

> boxplot(PFL.old, SHI.new, names=c("PFL.old","SHI.new"))

> qqnorm(SHI.new)
```

```
> qqline(SHI.new)

> x <- seq(0., 100., 1.0)

> qqplot(qnorm(ppoints(200)), x)

> qqline(x)

> qqplot(qnorm(ppoints(200)),SHI.new)

> qqline(SHI.new)

> qqplot(rnorm(1000),SHI.new)

> qqline(SHI.new)

> d1 <- rnorm(10000)

> d2 <- rnorm(10000)

> qqplot(d1,d1)

> qqline(d1)

> ### Empirical Cumulative Distribution Function

> plot(ecdf(SHI.new), do.points=FALSE)

> plot(ecdf(rnorm(1000, 45, 10)), do.points=FALSE, main="Norm Dist vs. SHI.new ECDF")

> lines(ecdf(SHI.new))

> plot(ecdf(SHI.old), do.points=FALSE, main="SHI.old vs. SHI.new ECDF")

> lines(ecdf(SHI.new))

> #fitting other distributions

> #Chi square

> qqplot(qchisq(ppoints(200), df=5),SHI.new)

> qqline(SHI.new)

> #Variable 2: AIR

> epi.results$AIR.new

 [1] 15.8 36.5 46.1 19.9 77.8 47.6 30.9 81.0 61.5 38.2 76.4 32.7  6.3 85.8 51.3 64.8
```

```
 [17] 41.3 25.0 16.8 23.0 22.9 34.5 36.2 68.7 32.0 39.8  9.1 63.1 18.2 17.0 72.3 15.6

 [33] 33.6 29.2 14.3 39.6 46.8 49.6 37.0 40.6 52.5 55.9 50.4  8.2 70.9 29.2 62.0 46.3

 [49] 37.3 31.7 22.9 32.4 26.9 60.9 22.5 25.4 61.7 82.2 65.2 30.6 46.9 34.5 66.9 32.4

 [65] 53.8 74.4 20.1 36.8 42.9 65.4 25.0 18.7 38.7 89.7  6.8 22.8 36.9 27.7 76.8 53.7

 [81] 52.3 35.8 59.9 42.9 44.5 25.0 54.5 41.5 30.6 13.7 45.1 40.0 11.8 43.2 53.2 67.1

 [97] 30.8 20.0 43.2 47.8 44.7 69.9 63.4 53.5 75.8 29.7 64.8 34.5 17.8 35.3 44.7 26.9

[113]  9.1 27.4  6.2 67.4 83.1 36.7 36.5 18.0 29.9 82.9 47.5  6.4 57.5 40.0 33.6 26.6

[129] 22.8 38.5 61.1 42.4 12.2 39.4 50.5  8.5 73.7 74.1 57.9 40.9 34.7 49.5 31.6 80.2

[145] 38.2 53.6 50.6 45.8 60.1 20.4 44.5 56.2 20.1 35.3 68.3 81.2 67.5 40.1 17.4 23.9

[161] 25.5 38.3 26.8 60.3 87.2 44.3 34.8 54.1 14.0 40.0 46.2 69.9 65.8 52.3 27.5 57.5

[177] 54.2 15.5 16.7 22.8
> epi.results[1,5]
[1] 30.7
> attach(epi.results)
The following objects are masked from epi.results (pos = 3):


  AGR.new, AGR.old, AIR.new, AIR.old, APO.new, APO.old, BCA.new, BCA.old,

  BDH.new, BDH.old, BER.new, BER.old, BTO.new, BTO.old, BTZ.new, BTZ.old,

  CBP.new, CBP.old, CCH.new, CCH.old, CDA.new, CDA.old, CDF.new, CDF.old,

  CHA.new, CHA.old, code, COE.new, COE.old, country, ECO.new, ECO.old,

  ECS.new, ECS.old, EPI.new, EPI.old, FCD.new, FCD.old, FCL.new, FCL.old,

  FGA.new, FGA.old, FLI.new, FLI.old, FSH.new, FSH.old, FSS.new, FSS.old,

  GHN.new, GHN.old, GTI.new, GTI.old, GTP.new, GTP.old, H2O.new, H2O.old,

  HFD.new, HFD.old, HLT.new, HLT.old, HMT.new, HMT.old, HPE.new, HPE.old,

  IFL.new, IFL.old, iso, LED.new, LED.old, LUF.new, LUF.old, MHP.new,
```

MHP.old, MKP.new, MKP.old, MPE.new, MPE.old, NDA.new, NDA.old, NOD.new,

NOD.old, NXA.new, NXA.old, OEB.new, OEB.old, OEC.new, OEC.old, OZD.new,

OZD.old, PAE.new, PAE.old, PAR.new, PAR.old, PCC.new, PCC.old, PFL.new,

PFL.old, PHL.new, PHL.old, PRS.new, PRS.old, PSU.new, PSU.old, RCY.new,

RCY.old, RLI.new, RLI.old, RMS.new, RMS.old, SDA.new, SDA.old, SHI.new,

SHI.old, SMW.new, SMW.old, SNM.new, SNM.old, SOE.new, SOE.old, SPI.new,

SPI.old, TBN.new, TBN.old, TCG.new, TCG.old, TKP.new, TKP.old, USD.new,

USD.old, UWD.new, UWD.old, VOE.new, VOE.old, WMG.new, WMG.old, WPC.new,

WPC.old, WRR.new, WRR.old, WRS.new, WRS.old, WWC.new, WWC.old, WWG.new,

WWG.old, WWR.new, WWR.old, WWT.new, WWT.old

The following objects are masked from epi.results.sub:

country, EPI.new, EPI.old

The following objects are masked from epi.results (pos = 5):

AGR.new, AGR.old, AIR.new, AIR.old, APO.new, APO.old, BCA.new, BCA.old,

BDH.new, BDH.old, BER.new, BER.old, BTO.new, BTO.old, BTZ.new, BTZ.old,

CBP.new, CBP.old, CCH.new, CCH.old, CDA.new, CDA.old, CDF.new, CDF.old,

CHA.new, CHA.old, code, COE.new, COE.old, country, ECO.new, ECO.old,

ECS.new, ECS.old, EPI.new, EPI.old, FCD.new, FCD.old, FCL.new, FCL.old,

FGA.new, FGA.old, FLI.new, FLI.old, FSH.new, FSH.old, FSS.new, FSS.old,

GHN.new, GHN.old, GTI.new, GTI.old, GTP.new, GTP.old, H2O.new, H2O.old,

HFD.new, HFD.old, HLT.new, HLT.old, HMT.new, HMT.old, HPE.new, HPE.old,

IFL.new, IFL.old, iso, LED.new, LED.old, LUF.new, LUF.old, MHP.new,

MHP.old, MKP.new, MKP.old, MPE.new, MPE.old, NDA.new, NDA.old, NOD.new,

NOD.old, NXA.new, NXA.old, OEB.new, OEB.old, OEC.new, OEC.old, OZD.new,

OZD.old, PAE.new, PAE.old, PAR.new, PAR.old, PCC.new, PCC.old, PFL.new,

PFL.old, PHL.new, PHL.old, PRS.new, PRS.old, PSU.new, PSU.old, RCY.new,

RCY.old, RLI.new, RLI.old, RMS.new, RMS.old, SDA.new, SDA.old, SHI.new,

SHI.old, SMW.new, SMW.old, SNM.new, SNM.old, SOE.new, SOE.old, SPI.new,

SPI.old, TBN.new, TBN.old, TCG.new, TCG.old, TKP.new, TKP.old, USD.new,

USD.old, UWD.new, UWD.old, VOE.new, VOE.old, WMG.new, WMG.old, WPC.new,

WPC.old, WRR.new, WRR.old, WRS.new, WRS.old, WWC.new, WWC.old, WWG.new,

WWG.old, WWR.new, WWR.old, WWT.new, WWT.old

> AIR.new

 [1] 15.8 36.5 46.1 19.9 77.8 47.6 30.9 81.0 61.5 38.2 76.4 32.7 6.3 85.8 51.3 64.8

[17] 41.3 25.0 16.8 23.0 22.9 34.5 36.2 68.7 32.0 39.8 9.1 63.1 18.2 17.0 72.3 15.6

[33] 33.6 29.2 14.3 39.6 46.8 49.6 37.0 40.6 52.5 55.9 50.4 8.2 70.9 29.2 62.0 46.3

[49] 37.3 31.7 22.9 32.4 26.9 60.9 22.5 25.4 61.7 82.2 65.2 30.6 46.9 34.5 66.9 32.4

[65] 53.8 74.4 20.1 36.8 42.9 65.4 25.0 18.7 38.7 89.7 6.8 22.8 36.9 27.7 76.8 53.7

[81] 52.3 35.8 59.9 42.9 44.5 25.0 54.5 41.5 30.6 13.7 45.1 40.0 11.8 43.2 53.2 67.1

[97] 30.8 20.0 43.2 47.8 44.7 69.9 63.4 53.5 75.8 29.7 64.8 34.5 17.8 35.3 44.7 26.9

[113] 9.1 27.4 6.2 67.4 83.1 36.7 36.5 18.0 29.9 82.9 47.5 6.4 57.5 40.0 33.6 26.6

[129] 22.8 38.5 61.1 42.4 12.2 39.4 50.5 8.5 73.7 74.1 57.9 40.9 34.7 49.5 31.6 80.2

[145] 38.2 53.6 50.6 45.8 60.1 20.4 44.5 56.2 20.1 35.3 68.3 81.2 67.5 40.1 17.4 23.9

[161] 25.5 38.3 26.8 60.3 87.2 44.3 34.8 54.1 14.0 40.0 46.2 69.9 65.8 52.3 27.5 57.5

[177] 54.2 15.5 16.7 22.8

```
> AIR.new[1]

[1] 15.8

> ## NA values

> na.indices <- is.na(AIR.new)

> ## drop NAs

> Air.new.compl <- AIR.new[!na.indices]

> ## convert to data frame and add country

> Air.new.compl <- data.frame(Country = country[!na.indices], AIR = AIR.new[!na.indices])

> ## summary stats

> summary(AIR.new)

  Min. 1st Qu.  Median    Mean 3rd Qu.   Max.

  6.20   26.90   40.00   42.29   55.98   89.70

> fivenum(AIR.new,na.rm=TRUE)

[1]  6.20 26.90 40.00 56.05 89.70

> ## histograms

> hist(AIR.new)

> hist(AIR.new, seq(0., 100., 2.0), prob=TRUE)

> rug(AIR.new)

> lines(density(AIR.new,na.rm=TRUE,bw=1))

> lines(density(AIR.new,na.rm=TRUE,bw="SJ"))

> x <- seq(0., 100., 1.0)

> qn<- dnorm(x,mean=42, sd=5,log=FALSE)

> lines(x,0.4*qn)

> qn<- dnorm(x,mean=65, sd=5,log=FALSE)

> lines(x,0.12*qn)
```

```
> boxplot(WWR.old, AIR.new, names=c("WWR.old","AIR.new"))

> qqnorm(AIR.new)

> qqline(AIR.new)

> x <- seq(0., 100., 1.0)

> qqplot(qnorm(ppoints(200)), x)

> qqline(x)

> qqplot(qnorm(ppoints(200)),AIR.new)

> qqline(AIR.new)

> qqplot(rnorm(1000),AIR.new)

> qqline(AIR.new)

> d1 <- rnorm(10000)

> d2 <- rnorm(10000)

> qqplot(d1,d1)

> qqline(d1)

> ### Empirical Cumulative Distribution Function

> plot(ecdf(AIR.new), do.points=FALSE)

> plot(ecdf(rnorm(1000, 45, 10)), do.points=FALSE, main="Norm Dist vs. AIR.new ECDF")

> lines(ecdf(AIR.new))

> plot(ecdf(AIR.old), do.points=FALSE, main="AIR.old vs. AIR.new ECDF")

> lines(ecdf(AIR.new))

> #fitting other distributions

> #Chi square

> qqplot(qchisq(ppoints(200), df=5),AIR.new)

> qqline(AIR.new)

> #Variable 3:USD
```

```
> epi.results$USD.new
  [1]  31.0  73.2  68.8  23.4  56.0  70.2  68.7  91.6  87.7  45.2  68.7  69.9  30.9
 [14]  61.1  71.9  87.4  51.4  18.9  27.6  44.7  77.9  18.4  57.6  88.0  87.8  14.7
 [27]  15.9  34.2  38.9  21.8  88.0   8.4   4.0  77.6  71.6  62.1  24.1  73.3  23.1
 [40]  91.1  59.8  87.5  89.6  24.1  91.1  24.4  53.0  42.8  62.7  50.3  48.7  36.3
 [53]  14.9  70.4  17.3  17.2  37.5  93.5  86.7  31.9  23.6  66.6  97.7  26.6  85.3
 [66]  55.7  31.8  19.0  19.0  43.5  17.4  36.1  76.9  93.7  25.3  36.6  61.5  58.2
 [79]  82.0  77.9  95.6  53.4  80.5  65.6  68.4  20.1  21.0  81.5  54.8  30.9  73.8
 [92]  73.2   8.9  18.0  70.4  93.2  12.1  16.0  57.7  61.3  17.3  88.9  36.2  25.7
[105]  61.2  59.6  39.3  59.0  58.3 100.0  52.6  19.3  34.2  19.2  33.0  87.7  85.6
[118]  44.7  11.4  14.0  71.0  97.4  78.4  28.2  48.4  20.2  55.1  52.9  45.5  83.9
[131]  92.5  80.6  22.9  65.9  68.3  23.2  54.8  51.9  64.4  36.1  69.2  25.5  83.9
[144]  58.0  16.9 100.0 100.0  95.7  31.7  24.9  92.7  88.6  55.6  36.4  46.5  96.0
[157]  98.2  72.1  28.1  19.4  61.2  31.5  17.1  50.9  63.0  64.3  66.9  54.2  22.1
[170]  72.2  80.6 100.0  90.9  75.6  65.4  31.2  49.3  52.2  20.1  19.2
> epi.results[1,5]
[1] 30.7
> attach(epi.results)
The following objects are masked from epi.results (pos = 3):

  AGR.new, AGR.old, AIR.new, AIR.old, APO.new, APO.old, BCA.new, BCA.old,
  BDH.new, BDH.old, BER.new, BER.old, BTO.new, BTO.old, BTZ.new, BTZ.old,
  CBP.new, CBP.old, CCH.new, CCH.old, CDA.new, CDA.old, CDF.new, CDF.old,
  CHA.new, CHA.old, code, COE.new, COE.old, country, ECO.new, ECO.old,
  ECS.new, ECS.old, EPI.new, EPI.old, FCD.new, FCD.old, FCL.new, FCL.old,
```

FGA.new, FGA.old, FLI.new, FLI.old, FSH.new, FSH.old, FSS.new, FSS.old,

GHN.new, GHN.old, GTI.new, GTI.old, GTP.new, GTP.old, H2O.new, H2O.old,

HFD.new, HFD.old, HLT.new, HLT.old, HMT.new, HMT.old, HPE.new, HPE.old,

IFL.new, IFL.old, iso, LED.new, LED.old, LUF.new, LUF.old, MHP.new,

MHP.old, MKP.new, MKP.old, MPE.new, MPE.old, NDA.new, NDA.old, NOD.new,

NOD.old, NXA.new, NXA.old, OEB.new, OEB.old, OEC.new, OEC.old, OZD.new,

OZD.old, PAE.new, PAE.old, PAR.new, PAR.old, PCC.new, PCC.old, PFL.new,

PFL.old, PHL.new, PHL.old, PRS.new, PRS.old, PSU.new, PSU.old, RCY.new,

RCY.old, RLI.new, RLI.old, RMS.new, RMS.old, SDA.new, SDA.old, SHI.new,

SHI.old, SMW.new, SMW.old, SNM.new, SNM.old, SOE.new, SOE.old, SPI.new,

SPI.old, TBN.new, TBN.old, TCG.new, TCG.old, TKP.new, TKP.old, USD.new,

USD.old, UWD.new, UWD.old, VOE.new, VOE.old, WMG.new, WMG.old, WPC.new,

WPC.old, WRR.new, WRR.old, WRS.new, WRS.old, WWC.new, WWC.old, WWG.new,

WWG.old, WWR.new, WWR.old, WWT.new, WWT.old

The following objects are masked from epi.results (pos = 4):

AGR.new, AGR.old, AIR.new, AIR.old, APO.new, APO.old, BCA.new, BCA.old,

BDH.new, BDH.old, BER.new, BER.old, BTO.new, BTO.old, BTZ.new, BTZ.old,

CBP.new, CBP.old, CCH.new, CCH.old, CDA.new, CDA.old, CDF.new, CDF.old,

CHA.new, CHA.old, code, COE.new, COE.old, country, ECO.new, ECO.old,

ECS.new, ECS.old, EPI.new, EPI.old, FCD.new, FCD.old, FCL.new, FCL.old,

FGA.new, FGA.old, FLI.new, FLI.old, FSH.new, FSH.old, FSS.new, FSS.old,

GHN.new, GHN.old, GTI.new, GTI.old, GTP.new, GTP.old, H2O.new, H2O.old,

HFD.new, HFD.old, HLT.new, HLT.old, HMT.new, HMT.old, HPE.new, HPE.old,

IFL.new, IFL.old, iso, LED.new, LED.old, LUF.new, LUF.old, MHP.new,

MHP.old, MKP.new, MKP.old, MPE.new, MPE.old, NDA.new, NDA.old, NOD.new,

NOD.old, NXA.new, NXA.old, OEB.new, OEB.old, OEC.new, OEC.old, OZD.new,

OZD.old, PAE.new, PAE.old, PAR.new, PAR.old, PCC.new, PCC.old, PFL.new,

PFL.old, PHL.new, PHL.old, PRS.new, PRS.old, PSU.new, PSU.old, RCY.new,

RCY.old, RLI.new, RLI.old, RMS.new, RMS.old, SDA.new, SDA.old, SHI.new,

SHI.old, SMW.new, SMW.old, SNM.new, SNM.old, SOE.new, SOE.old, SPI.new,

SPI.old, TBN.new, TBN.old, TCG.new, TCG.old, TKP.new, TKP.old, USD.new,

USD.old, UWD.new, UWD.old, VOE.new, VOE.old, WMG.new, WMG.old, WPC.new,

WPC.old, WRR.new, WRR.old, WRS.new, WRS.old, WWC.new, WWC.old, WWG.new,

WWG.old, WWR.new, WWR.old, WWT.new, WWT.old

The following objects are masked from epi.results.sub:

country, EPI.new, EPI.old

The following objects are masked from epi.results (pos = 6):

AGR.new, AGR.old, AIR.new, AIR.old, APO.new, APO.old, BCA.new, BCA.old,

BDH.new, BDH.old, BER.new, BER.old, BTO.new, BTO.old, BTZ.new, BTZ.old,

CBP.new, CBP.old, CCH.new, CCH.old, CDA.new, CDA.old, CDF.new, CDF.old,

CHA.new, CHA.old, code, COE.new, COE.old, country, ECO.new, ECO.old,

ECS.new, ECS.old, EPI.new, EPI.old, FCD.new, FCD.old, FCL.new, FCL.old,

FGA.new, FGA.old, FLI.new, FLI.old, FSH.new, FSH.old, FSS.new, FSS.old,

GHN.new, GHN.old, GTI.new, GTI.old, GTP.new, GTP.old, H2O.new, H2O.old,

HFD.new, HFD.old, HLT.new, HLT.old, HMT.new, HMT.old, HPE.new, HPE.old,

IFL.new, IFL.old, iso, LED.new, LED.old, LUF.new, LUF.old, MHP.new,

MHP.old, MKP.new, MKP.old, MPE.new, MPE.old, NDA.new, NDA.old, NOD.new,

NOD.old, NXA.new, NXA.old, OEB.new, OEB.old, OEC.new, OEC.old, OZD.new,

OZD.old, PAE.new, PAE.old, PAR.new, PAR.old, PCC.new, PCC.old, PFL.new,

PFL.old, PHL.new, PHL.old, PRS.new, PRS.old, PSU.new, PSU.old, RCY.new,

RCY.old, RLI.new, RLI.old, RMS.new, RMS.old, SDA.new, SDA.old, SHI.new,

SHI.old, SMW.new, SMW.old, SNM.new, SNM.old, SOE.new, SOE.old, SPI.new,

SPI.old, TBN.new, TBN.old, TCG.new, TCG.old, TKP.new, TKP.old, USD.new,

USD.old, UWD.new, UWD.old, VOE.new, VOE.old, WMG.new, WMG.old, WPC.new,

WPC.old, WRR.new, WRR.old, WRS.new, WRS.old, WWC.new, WWC.old, WWG.new,

WWG.old, WWR.new, WWR.old, WWT.new, WWT.old

> USD.new

```
  [1] 31.0 73.2 68.8 23.4 56.0 70.2 68.7 91.6 87.7 45.2 68.7 69.9 30.9
 [14] 61.1 71.9 87.4 51.4 18.9 27.6 44.7 77.9 18.4 57.6 88.0 87.8 14.7
 [27] 15.9 34.2 38.9 21.8 88.0  8.4  4.0 77.6 71.6 62.1 24.1 73.3 23.1
 [40] 91.1 59.8 87.5 89.6 24.1 91.1 24.4 53.0 42.8 62.7 50.3 48.7 36.3
 [53] 14.9 70.4 17.3 17.2 37.5 93.5 86.7 31.9 23.6 66.6 97.7 26.6 85.3
 [66] 55.7 31.8 19.0 19.0 43.5 17.4 36.1 76.9 93.7 25.3 36.6 61.5 58.2
 [79] 82.0 77.9 95.6 53.4 80.5 65.6 68.4 20.1 21.0 81.5 54.8 30.9 73.8
 [92] 73.2  8.9 18.0 70.4 93.2 12.1 16.0 57.7 61.3 17.3 88.9 36.2 25.7
[105] 61.2 59.6 39.3 59.0 58.3 100.0 52.6 19.3 34.2 19.2 33.0 87.7 85.6
[118] 44.7 11.4 14.0 71.0 97.4 78.4 28.2 48.4 20.2 55.1 52.9 45.5 83.9
[131] 92.5 80.6 22.9 65.9 68.3 23.2 54.8 51.9 64.4 36.1 69.2 25.5 83.9
```

```
[144] 58.0 16.9 100.0 100.0 95.7 31.7 24.9 92.7 88.6 55.6 36.4 46.5 96.0

[157] 98.2 72.1 28.1 19.4 61.2 31.5 17.1 50.9 63.0 64.3 66.9 54.2 22.1

[170] 72.2 80.6 100.0 90.9 75.6 65.4 31.2 49.3 52.2 20.1 19.2

> USD.new[1]

[1] 31

> ## NA values

> na.indices <- is.na(USD.new)

> ## drop NAs

> Usd.new.compl <- USD.new[!na.indices]

> ## convert to data frame and add country

> Usd.new.compl <- data.frame(Country = country[!na.indices], USD = USD.new[!na.indices])

> ## summary stats

> summary(USD.new)

  Min. 1st Qu.  Median   Mean 3rd Qu.   Max.

  4.00  27.98  55.35  53.74  74.25 100.00

> fivenum(USD.new,na.rm=TRUE)

[1]  4.00  27.85  55.35  74.70 100.00

> ## histograms

> hist(USD.new)

> hist(USD.new, seq(0., 100., 2.0), prob=TRUE)

> rug(USD.new)

> lines(density(USD.new,na.rm=TRUE,bw=1))

> lines(density(USD.new,na.rm=TRUE,bw="SJ"))

> x <- seq(0., 100., 1.0)

> qn<- dnorm(x,mean=42, sd=5,log=FALSE)
```

```
> lines(x,0.4*qn)

> qn<- dnorm(x,mean=65, sd=5,log=FALSE)

> lines(x,0.12*qn)

> boxplot(USD.new, H2O.new, names=c("USD.new","H2O.new"))

> qqnorm(USD.new)

> qqline(USD.new)

> x <- seq(0., 100., 1.0)

> qqplot(qnorm(ppoints(200)), x)

> qqline(x)

> qqplot(qnorm(ppoints(200)),USD.new)

> qqline(USD.new)

> qqplot(rnorm(1000),USD.new)

> qqline(USD.new)

> d1 <- rnorm(10000)

> d2 <- rnorm(10000)

> qqplot(d1,d1)

> qqline(d1)

> ### Empirical Cumulative Distribution Function

> plot(ecdf(USD.new), do.points=FALSE)

> plot(ecdf(rnorm(1000, 45, 10)), do.points=FALSE, main="Norm Dist vs. USD.new ECDF")

> lines(ecdf(USD.new))

> plot(ecdf(USD.old), do.points=FALSE, main="USD.old vs. USD.new ECDF")

> lines(ecdf(USD.new))

> #fitting other distributions

> #Chi square
```

```
> qqplot(qchisq(ppoints(200), df=5),USD.new)

> qqline(USD.new)

> #Boxplot comparing 3 variables

> boxplot(TCG.old, AIR.new, HFD.old, names=c("WWR.old","AIR.new", "HFD.old"))

> #Q-Q plots for 3 variables compared to some known distribution

> par(mfrow = c(1, 3))

> qqnorm(AIR.new, main = "Q-Q Plot of AIR.new")

> qqline(AIR.new, col = "blue")

> qqnorm(HPE.new, main = "Q-Q Plot of HPE.new")

> qqline(HPE.new, col = "orange")

> qqnorm(OZD.new, main = "Q-Q Plot of OZD.new")

> qqline(OZD.new, col = "purple")

> par(mfrow = c(1, 1))

> #ECDF plots for 3 variables compared to each other

> plot(ecdf(AIR.new), do.points=FALSE, main="AIR.new vs. SHI.new vs. HPE.new ECDF")

> lines(ecdf(epi.results$AIR.new))

> lines(ecdf(epi.results$SHI.new))

> lines(ecdf(epi.results$HPE.new))

> #Exercise 2: linear models

> #Variable 1: GTI

> ## drop country populations that don't exist in epi results

> populations <- populations_2023[-which(!populations_2023$Country %in% epi.results$country),]

> ## sort populations by country name

> populations <- populations[order(populations$Country),]

> ## drop country results that don't exist in populations
```

```
> epi.results.sub <- epi.results[-which(!epi.results$country %in% populations$Country),]

> ## sort results by country name

> epi.results.sub <- epi.results.sub[order(epi.results.sub$country),]

> ## only keep relevant columns

> epi.results.sub <- epi.results.sub[,c("country","GTI.old","GTI.new")]

> ## convert to mnumeric

> epi.results.sub$population <- as.numeric(populations$Population)

> ## compute population log

> epi.results.sub$population_log <- log10(epi.results.sub$population)

> boxplot(epi.results.sub$population_log)

> attach(epi.results.sub)
The following objects are masked from epi.results (pos = 3):


    country, GTI.new, GTI.old


The following objects are masked from epi.results (pos = 4):


    country, GTI.new, GTI.old


The following objects are masked from epi.results (pos = 5):


    country, GTI.new, GTI.old


The following objects are masked from epi.results.sub (pos = 6):
```

country, population, population_log

The following objects are masked from epi.results (pos = 7):

country, GTI.new, GTI.old

```
> ## created linear model of EPI.new = a(population_log) + b
> lin.mod.gtinew <- lm(GTI.new~population_log,epi.results.sub)
> plot(GTI.new~population_log)
> abline(lin.mod.gtinew)
> summary(lin.mod.gtinew)
```

Call:
lm(formula = GTI.new ~ population_log, data = epi.results.sub)

Residuals:

Min	1Q	Median	3Q	Max
-38.972	-8.386	0.561	9.289	39.576

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	44.7304	9.4774	4.720	5.06e-06	***
population_log	-0.8811	1.3573	-0.649	0.517	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14.47 on 163 degrees of freedom

Multiple R-squared: 0.002579,   Adjusted R-squared: -0.003541

F-statistic: 0.4214 on 1 and 163 DF, p-value: 0.5172

```
> plot(lin.mod.gtinew)

Hit <Return> to see next plot: ggplot(epi.results.sub, aes(x = population_log, y = GTI.new)) +

Hit <Return> to see next plot:   geom_point() +

Hit <Return> to see next plot:   stat_smooth(method = "lm")

Hit <Return> to see next plot: ggplot(lin.mod.gtinew, aes(x = .fitted, y = .resid)) +

>  geom_point() +

+  geom_hline(yintercept = 0) +

+  labs(title='Residual vs. Fitted Values Plot', x='Fitted Values', y='Residuals')
```

Error in `+.gg`:

! Cannot add <ggproto> objects together.

i Did you forget to add this object to a <ggplot> object?

Run `rlang::last_trace()` to see where the error occurred.

```
> ggplot(lin.mod.gtinew, aes(x = .fitted, y = .resid)) +

+  geom_point() +

+  geom_hline(yintercept = 0) +

+  labs(title='Residual vs. Fitted Values Plot', x='Fitted Values', y='Residuals')

> ## another lm

> lin.mod.pop <- lm(population_log~GTI.new,epi.results.sub)

> plot(population_log~GTI.new)

> abline(lin.mod.pop)
```

```
> summary(lin.mod.pop)


Call:

lm(formula = population_log ~ GTI.new, data = epi.results.sub)


Residuals:

    Min     1Q  Median     3Q     Max

-2.35789 -0.48289  0.07572  0.55159  2.20488


Coefficients:

         Estimate Std. Error t value Pr(>|t|)

(Intercept)  7.046235   0.185833  37.917   <2e-16 ***

GTI.new    -0.002927   0.004508  -0.649   0.517

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Residual standard error: 0.8339 on 163 degrees of freedom

Multiple R-squared:  0.002579,     Adjusted R-squared:  -0.003541

F-statistic: 0.4214 on 1 and 163 DF,  p-value: 0.5172


> plot(lin.mod.pop)

Hit <Return> to see next plot: ggplot(epi.results.sub, aes(x = GTI.new, y = population_log)) +

Hit <Return> to see next plot:   geom_point() +

Hit <Return> to see next plot:   stat_smooth(method = "lm")

Hit <Return> to see next plot: ggplot(lin.mod.pop, aes(x = .fitted, y = .resid)) +
```

```
>   geom_point() +

+   geom_hline(yintercept = 0) +

+   labs(title='Residual vs. Fitted Values Plot', x='Fitted Values', y='Residuals')
```

Error in `+.gg`:

! Cannot add <ggproto> objects together.

ℹ Did you forget to add this object to a <ggplot> object?

Run `rlang::last_trace()` to see where the error occurred.

```
> ggplot(lin.mod.pop, aes(x = .fitted, y = .resid)) +

+   geom_point() +

+   geom_hline(yintercept = 0) +

+   labs(title='Residual vs. Fitted Values Plot', x='Fitted Values', y='Residuals')

> #Variable 2: GTP

> ## drop country populations that don't exist in epi results

> populations <- populations_2023[-which(!populations_2023$Country %in% epi.results$country),]

> ## sort populations by country name

> populations <- populations[order(populations$Country),]

> ## drop country results that don't exist in populations

> epi.results.sub <- epi.results[-which(!epi.results$country %in% populations$Country),]

> ## sort results by country name

> epi.results.sub <- epi.results.sub[order(epi.results.sub$country),]

> ## only keep relevant columns

> epi.results.sub <- epi.results.sub[,c("country","GTP.old","GTP.new")]

> ## convert to mnumeric

> epi.results.sub$population <- as.numeric(populations$Population)

> ## compute population log
```

```
> epi.results.sub$population_log <- log10(epi.results.sub$population)

> boxplot(epi.results.sub$population_log)

> attach(epi.results.sub)
```

The following objects are masked from epi.results.sub (pos = 3):

    country, population, population_log

The following objects are masked from epi.results (pos = 4):

    country, GTP.new, GTP.old

The following objects are masked from epi.results (pos = 5):

    country, GTP.new, GTP.old

The following objects are masked from epi.results (pos = 6):

    country, GTP.new, GTP.old

The following objects are masked from epi.results.sub (pos = 7):

    country, population, population_log

The following objects are masked from epi.results (pos = 8):

country, GTP.new, GTP.old

> ## created linear model of EPI.new = a(population_log) + b

> lin.mod.gtpnew <- lm(GTP.new~population_log,epi.results.sub)

> plot(GTP.new~population_log)

> abline(lin.mod.gtpnew)

> summary(lin.mod.gtpnew)

Call:

lm(formula = GTP.new ~ population_log, data = epi.results.sub)

Residuals:

   Min   1Q Median   3Q   Max

-39.205 -5.941   1.612   7.591 30.229

Coefficients:

          Estimate Std. Error t value Pr(>|t|)

(Intercept)   42.377   7.957   5.326 3.29e-07 ***

population_log -0.490   1.139 -0.430   0.668

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 12.15 on 163 degrees of freedom

Multiple R-squared: 0.001133,   Adjusted R-squared: -0.004995

F-statistic: 0.1849 on 1 and 163 DF, p-value: 0.6678

```
> plot(lin.mod.gtpnew)

Hit <Return> to see next plot: ggplot(epi.results.sub, aes(x = population_log, y = GTP.new)) +

Hit <Return> to see next plot:   geom_point() +

Hit <Return> to see next plot:   stat_smooth(method = "lm")

Hit <Return> to see next plot: ggplot(lin.mod.gtpnew, aes(x = .fitted, y = .resid)) +

>  geom_point() +

+  geom_hline(yintercept = 0) +

+  labs(title='Residual vs. Fitted Values Plot', x='Fitted Values', y='Residuals')
```

Error in `+.gg`:

```
! Cannot add <ggproto> objects together.

i Did you forget to add this object to a <ggplot> object?

Run `rlang::last_trace()` to see where the error occurred.

> ggplot(lin.mod.gtpnew, aes(x = .fitted, y = .resid)) +

+  geom_point() +

+  geom_hline(yintercept = 0) +

+  labs(title='Residual vs. Fitted Values Plot', x='Fitted Values', y='Residuals')

> ## another lm

> lin.mod.pop <- lm(population_log~GTP.new,epi.results.sub)

> plot(population_log~GTP.new)

> abline(lin.mod.pop)

> summary(lin.mod.pop)


Call:

lm(formula = population_log ~ GTP.new, data = epi.results.sub)
```

Residuals:

```
   Min     1Q  Median    3Q    Max
-2.33755 -0.46745  0.05763  0.57412  2.21514
```

Coefficients:

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 7.023334  0.219457  32.00  <2e-16 ***
GTP.new    -0.002312  0.005378  -0.43   0.668
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 0.8345 on 163 degrees of freedom

Multiple R-squared: 0.001133,   Adjusted R-squared: -0.004995

F-statistic: 0.1849 on 1 and 163 DF, p-value: 0.6678

```
> plot(lin.mod.pop)
Hit <Return> to see next plot: ggplot(epi.results.sub, aes(x = GTP.new, y = population_log)) +
Hit <Return> to see next plot:   geom_point() +
Hit <Return> to see next plot:   stat_smooth(method = "lm")
Hit <Return> to see next plot: ggplot(lin.mod.pop, aes(x = .fitted, y = .resid)) +
>  geom_point() +
+  geom_hline(yintercept = 0) +
+  labs(title='Residual vs. Fitted Values Plot', x='Fitted Values', y='Residuals')
```
Error in `+.gg`:

! Cannot add <ggproto> objects together.

i Did you forget to add this object to a <ggplot> object?

Run `rlang::last_trace()` to see where the error occurred.

```
> ggplot(lin.mod.pop, aes(x = .fitted, y = .resid)) +
+   geom_point() +
+   geom_hline(yintercept = 0) +
+   labs(title='Residual vs. Fitted Values Plot', x='Fitted Values', y='Residuals')
> #Variable #3: PAR
> ## drop country populations that don't exist in epi results
> populations <- populations_2023[-which(!populations_2023$Country %in% epi.results$country),]
> ## sort populations by country name
> populations <- populations[order(populations$Country),]
> ## drop country results that don't exist in populations
> epi.results.sub <- epi.results[-which(!epi.results$country %in% populations$Country),]
> ## sort results by country name
> epi.results.sub <- epi.results.sub[order(epi.results.sub$country),]
> ## only keep relevant columns
> epi.results.sub <- epi.results.sub[,c("country","PAR.old","PAR.new")]
> ## convert to mnumeric
> epi.results.sub$population <- as.numeric(populations$Population)
> ## compute population log
> epi.results.sub$population_log <- log10(epi.results.sub$population)
> boxplot(epi.results.sub$population_log)
> attach(epi.results.sub)
```

The following objects are masked from epi.results.sub (pos = 3):

country, population, population_log

The following objects are masked from epi.results.sub (pos = 4):

country, population, population_log

The following objects are masked from epi.results (pos = 5):

country, PAR.new, PAR.old

The following objects are masked from epi.results (pos = 6):

country, PAR.new, PAR.old

The following objects are masked from epi.results (pos = 7):

country, PAR.new, PAR.old

The following objects are masked from epi.results.sub (pos = 8):

country, population, population_log

The following objects are masked from epi.results (pos = 9):

country, PAR.new, PAR.old

```
> ## created linear model of EPI.new = a(population_log) + b

> lin.mod.parnew <- lm(PAR.new~population_log,epi.results.sub)

> plot(PAR.new~population_log)

> abline(lin.mod.parnew)

> summary(lin.mod.parnew)


Call:
lm(formula = PAR.new ~ population_log, data = epi.results.sub)


Residuals:
   Min    1Q  Median    3Q    Max
-55.868 -18.302  -3.577  16.598  70.300


Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    112.001    22.018   5.087 1.09e-06 ***
population_log  -9.887     3.095  -3.194 0.00171 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Residual standard error: 26.49 on 148 degrees of freedom
 (15 observations deleted due to missingness)
Multiple R-squared:  0.0645,        Adjusted R-squared:  0.05818
```

F-statistic: 10.2 on 1 and 148 DF, p-value: 0.001712

```
> plot(lin.mod.parnew)

Hit <Return> to see next plot: ggplot(epi.results.sub, aes(x = population_log, y = PAR.new)) +

Hit <Return> to see next plot:   geom_point() +

Hit <Return> to see next plot:   stat_smooth(method = "lm")

Hit <Return> to see next plot: ggplot(lin.mod.parnew, aes(x = .fitted, y = .resid)) +

>   geom_point() +

+   geom_hline(yintercept = 0) +

+   labs(title='Residual vs. Fitted Values Plot', x='Fitted Values', y='Residuals')
```

Error in `+.gg`:

! Cannot add <ggproto> objects together.

i Did you forget to add this object to a <ggplot> object?

Run `rlang::last_trace()` to see where the error occurred.

```
> ggplot(lin.mod.parnew, aes(x = .fitted, y = .resid)) +

+   geom_point() +

+   geom_hline(yintercept = 0) +

+   labs(title='Residual vs. Fitted Values Plot', x='Fitted Values', y='Residuals')

> ## another lm

> lin.mod.pop <- lm(population_log~PAR.new,epi.results.sub)

> plot(population_log~PAR.old)

> abline(lin.mod.pop)

> summary(lin.mod.pop)
```

Call:

```
lm(formula = population_log ~ PAR.new, data = epi.results.sub)


Residuals:

   Min     1Q  Median     3Q    Max

-1.83706 -0.46733  0.02134  0.40147  1.80392


Coefficients:

        Estimate Std. Error t value Pr(>|t|)

(Intercept)  7.353857   0.102209  71.949  < 2e-16 ***

PAR.new     -0.006524   0.002042  -3.194  0.00171 **

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Residual standard error: 0.6804 on 148 degrees of freedom

 (15 observations deleted due to missingness)

Multiple R-squared:  0.0645,        Adjusted R-squared:  0.05818

F-statistic:  10.2 on 1 and 148 DF,  p-value: 0.001712


> plot(lin.mod.pop)

Hit <Return> to see next plot: ggplot(epi.results.sub, aes(x = PAR.old, y = population_log)) +

Hit <Return> to see next plot:   geom_point() +

Hit <Return> to see next plot:   stat_smooth(method = "lm")

Hit <Return> to see next plot: ggplot(lin.mod.pop, aes(x = .fitted, y = .resid)) +

>  geom_point() +

+  geom_hline(yintercept = 0) +
```

```
+   labs(title='Residual vs. Fitted Values Plot', x='Fitted Values', y='Residuals')
```

Error in `+.gg`:

! Cannot add <ggproto> objects together.

i Did you forget to add this object to a <ggplot> object?

Run `rlang::last_trace()` to see where the error occurred.

```
> ggplot(lin.mod.pop, aes(x = .fitted, y = .resid)) +
+   geom_point() +
+   geom_hline(yintercept = 0) +
+   labs(title='Residual vs. Fitted Values Plot', x='Fitted Values', y='Residuals')

>
```