WH(Week12)

student ID:106000229
Credit help:106000209

Question 1)

```
> names(auto) <- c("mpg", "cylinders", "displacement",
"horsepower", "weight", "acceleration", "model_year",
"origin", "car_name")
> car <- transform(auto, horsepower = as.numeric(horsepower))
> cars_log <- with(car, data.frame(log(mpg), log(cylinders),
log(displacement), log(horsepower), log(weight),
log(acceleration), model_year, origin))
> names(cars_log) <- c("<mpg>.log", "<cylinders>.log",
"<displacement>.log", "<horsepower>.log", "<weight>.log",
"<acceleration>.log", "model_year", "origin")
```

a.i

```
> lm_cars_log <- lm(mpg.log ~ cylinders.log + displacement.log
+ horsepower.log + weight.log + acceleration.log + model_year
+ factor(origin), data = cars_log)
> summary(lm_cars_log)
```

Call:
lm(formula = mpg.log ~ cylinders.log + displacement.log + horsepower.log +
    weight.log + acceleration.log + model_year + factor(origin),
    data = cars_log)

Residuals:

Min	1Q	Median	3Q	Max
-0.39728	-0.06892	0.00549	0.06297	0.38528

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	7.263274	0.361993	20.065	< 2e-16	***
cylinders.log	-0.086473	0.061078	-1.416	0.15765	
displacement.log	0.021415	0.058267	0.368	0.71342	
horsepower.log	-0.279597	0.057937	-4.826	2.02e-06	***
weight.log	-0.594565	0.085018	-6.993	1.21e-11	***
acceleration.log	-0.164113	0.059651	-2.751	0.00622	**
model_year	0.030428	0.001772	17.175	< 2e-16	***
factor(origin)2	0.051206	0.020884	2.452	0.01466	*
factor(origin)3	0.047401	0.020585	2.303	0.02183	*

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1128 on 382 degrees of freedom
(6 observations deleted due to missingness)
Multiple R-squared: 0.8925, Adjusted R-squared: 0.8902
F-statistic: 396.3 on 8 and 382 DF, p-value: < 2.2e-16

ANS: horsepower, weight, acceleration

a.ii
ANS: horsepower.log

a.iii
ANS:
opposite: horsepower.log, weight.log
insignificant: cylinder.log

b.i

```
> regr_wt <- lm(mpg ~ weight, data = car)
> summary(regr_wt)
```

Call:
lm(formula = mpg ~ weight, data = car)

Residuals:
```
    Min      1Q  Median      3Q     Max
-12.0136  -2.8081  -0.3488   2.1128  16.4790
```

Coefficients:
```
              Estimate  Std.  Error      t value    Pr(>|t|)
(Intercept) 46.3130164  0.7962549   58.16   <2e-16 ***
weight      -0.0076739    0.0002579  -29.75   <2e-16 ***
---
```
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.35 on 395 degrees of freedom
Multiple R-squared: 0.6915, Adjusted R-squared: 0.6907
F-statistic: 885.3 on 1 and 395 DF, p-value: < 2.2e-16

b.ii

```
> regr_wt_log <- lm(mpg.log ~ weight.log, data = cars_log)
> summary(regr_wt_log)
```

Call:

lm(formula = mpg.log ~ weight.log, data = cars_log)

Residuals:
```
    Min      1Q  Median      3Q     Max
```

-0.52408 -0.10441 -0.00805 0.10165 0.59384

Coefficients:

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

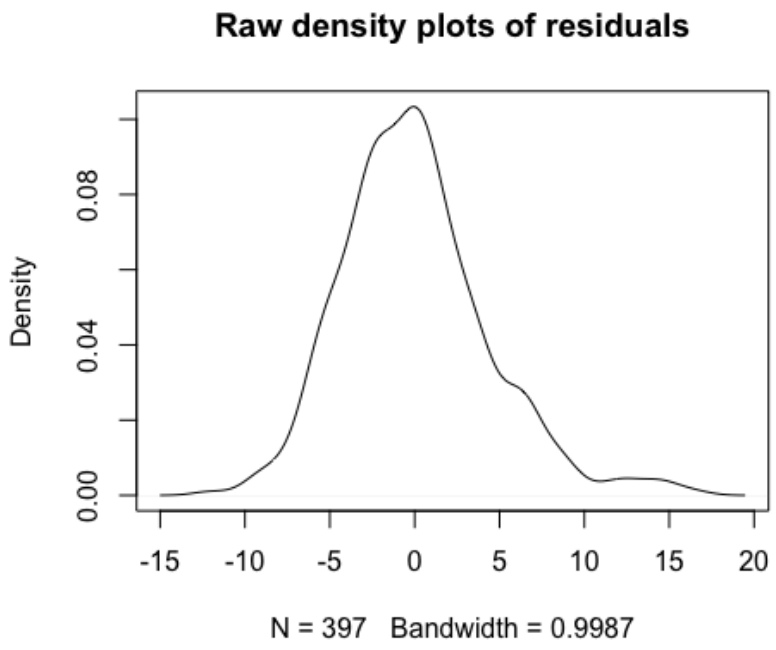
(Intercept) 11.5219 0.2349 49.06 <2e-16 *** weight.log -1.0583 0.0295 -35.87 <2e-16 ***

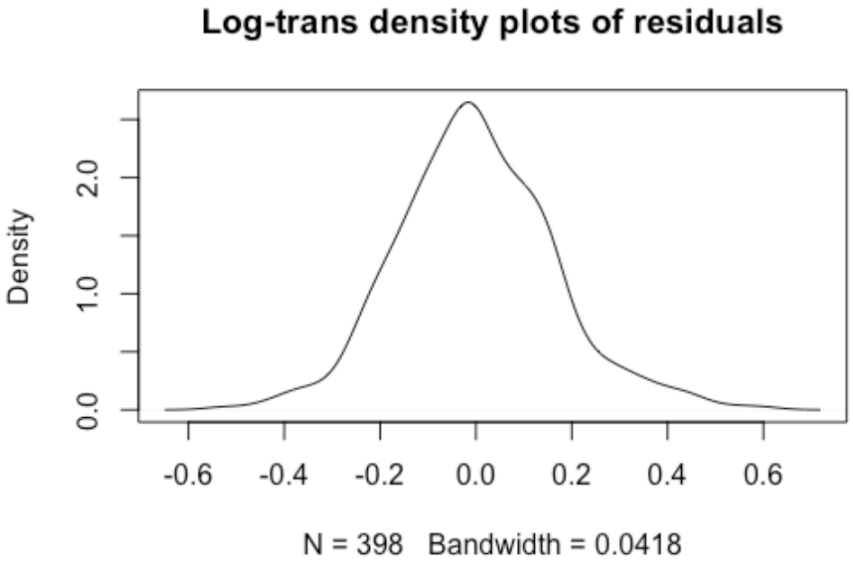
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.165 on 396 degrees of freedom Multiple R-squared: 0.7647, Adjusted R-squared: 0.7641 F-statistic: 1287 on 1 and 396 DF, p-value: < 2.2e-16

b.iii

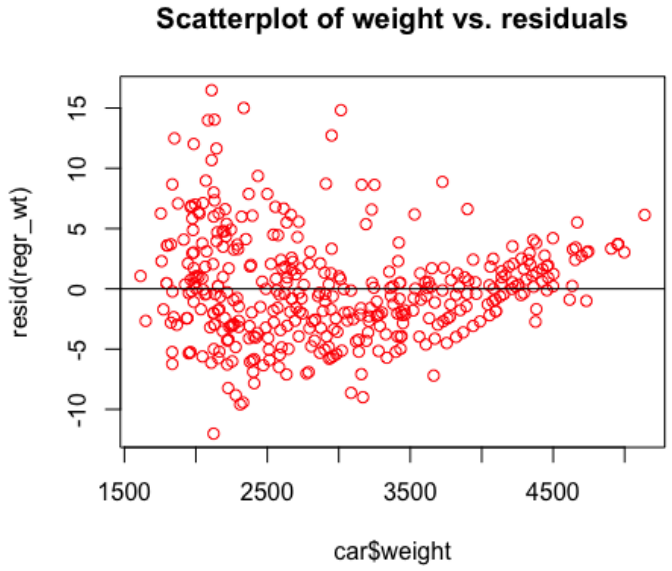
```
> plot(density(regr_wt$residuals), main="Raw density plots of residuals")
> plot(density(regr_wt_log$residuals), main="Log-trans density plots of residuals")
```





b.iii.2

```
> plot(car$weight,resid(regr_wt), col="red",lwd=1,
main="Scatterplot of weight vs. residuals")
> abline(h=0, col="black")
```




```
> plot(log(car$weight),resid(regr_wt_log), col="red",lwd=1,
main="Scatterplot of log.weight. vs. residuals")
> abline(h=0, col="black")
```

b.iv

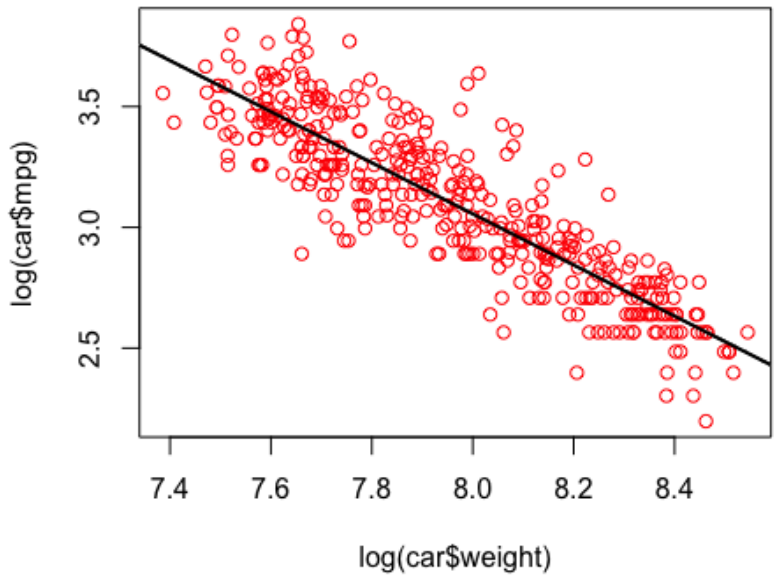
ANS: data has been log is better.

b.v

1% `log.weight` *leads to slope% increase/decrease in mpg*

c.i

```
> plot(log(car$weight), log(car$mpg), col=NA, pch=19)
> boot_regr <- function(model, dataset){
+     boot_index <- sample(1:nrow(dataset), replace=TRUE)
+     data_boot <- dataset[boot_index,]
+     regr_boot <- lm(model, data=data_boot)
+     abline(regr_boot, lwd=1, col=rgb(0.7, 0.7, 0.7, 0.5))
+     regr_boot$coefficients
+ }
> set.seed(42)
> coeffs <- replicate(3000, boot_regr(log(car$mpg)~log(car$weight), car))
> points(log(car$weight), log(car$mpg), col="red")
> abline(a=mean(coeffs["(Intercept)",]),
+        b=mean(coeffs["log(car$weight)",]), lwd=2)
```



Question 2)

a.

```
> regr_log <- lm(mpg.log ~ cylinders.log + displacement.log + horsepower.log
+ weight.log + acceleration.log + model_year +
+                     factor(origin), data=cars_log)
> weight_regr <- lm(weight.log ~ cylinders.log + displacement.log +
+                         horsepower.log +acceleration.log + model_year
+factor(origin), data = cars_log, na.action = na.exclude)
> r2_weight <- summary(weight_regr)$r.squared
> vif_weight <- 1/(1-r2_weight)
> sqrt(vif_weight)
```

[1] 4.189796

b.i

```
> vif(regr_log)
```

	GVIF	Df	GVIF^(1/(2*Df))
cylinders.log	10.427748	1	3.229202
displacement.log	29.533554	1	5.434478
horsepower.log	12.151794	1	3.485942
weight.log	17.554391	1	4.189796
acceleration.log	3.566907	1	1.888626
model_year	1.301139	1	1.140675
factor(origin)	2.653352	2	1.276288

b.iii

```
> regr_log_1 <- lm(mpg.log ~ cylinders.log + horsepower.log +
weight.log + acceleration.log + model_year +factor(origin),
data=cars_log)
> vif(regr_log_1)
```

	GVIF	Df	GVIF^(1/(2*Df))
cylinders.log	5.422425	1	2.328610
horsepower.log	12.133630	1	3.483336
weight.log	11.225309	1	3.350419
acceleration.log	3.324384	1	1.823289
model_year	1.289027	1	1.135353
factor(origin)	1.894961	2	1.173276

b.iv

```
> summary(regr_log_1)
```

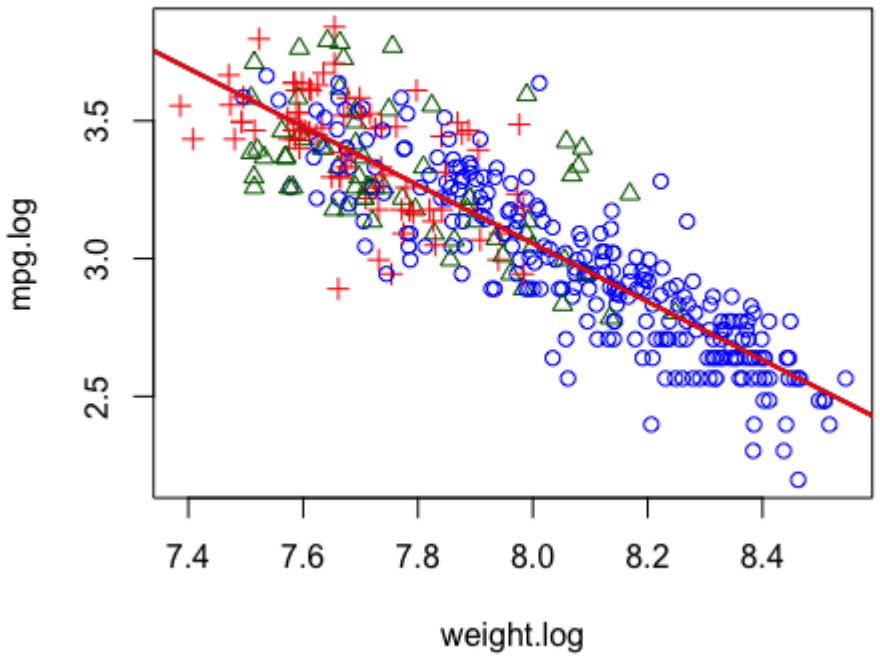
```
Call:
lm(formula = mpg.log ~ cylinders.log + horsepower.log + weight.log +
    acceleration.log + model_year + factor(origin), data = cars_log)

Residuals:
     Min       1Q   Median       3Q      Max
-0.40077 -0.06839  0.00569  0.06173  0.39110

Coefficients:
                    Estimate       Std.  Error t value    Pr(>|t|)
(Intercept)       7.223532   0.345076  20.933  < 2e-16 ***
cylinders.log    -0.070921   0.043994  -1.612  0.10777
horsepower.log   -0.280420   0.057828  -4.849 1.81e-06 ***
weight.log       -0.575802   0.067909  -8.479 4.98e-16 ***
acceleration.log -0.169830   0.057522  -2.952  0.00335 **
model_year        0.030365   0.001761  17.239  < 2e-16 ***
factor(origin)2   0.047477   0.018234   2.604  0.00958 **
factor(origin)3   0.043957   0.018307   2.401  0.01682 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1126 on 383 degrees of freedom
  (6 observations deleted due to missingness)
Multiple R-squared:  0.8924,  Adjusted R-squared:  0.8905
F-statistic:   454 on 7 and 383 DF,  p-value: < 2.2e-16
```

c.
ANS:
Yes, we lost some data, but it didn't hurt our explanation too much.

Question 3)

```
> origin_colors = c("#f2cc8f", "#81b29a", "#e07a5f")
> with(cars_log, plot(weight.log, mpg.log, pch=origin,
col=origin_colors[origin]))
```

```
> cars_us <- subset(cars_log, origin==1)
> wt_regr_us <- lm(mpg.log ~ weight.log, data=cars_us)
> abline(wt_regr_us, col=origin_colors[1], lwd=2)
####
> cars_eur <- subset(cars_log, origin==2)
> wt_regr_eur <- lm(mpg.log ~ weight.log, data=cars_eur) >
abline(wt_regr_eur, col=origin_colors[2], lwd=2)
####
> cars_jap <- subset(cars_log, origin==3)
> wt_regr_jap <- lm(mpg.log ~ weight.log, data=cars_jap) >
abline(wt_regr_jap, col=origin_colors[3], lwd=2)
```


mpg.log
2.5
3.0
3.5
7.4
7.6
7.8
8.0
8.2
8.4
weight.log