

Lab 5

Grace Randall

2024-03-05

Question 1

```
lm1 <- lm(formula = agb ~ 1, data=df)
summary(lm1)
```

```
##
## Call:
## lm(formula = agb ~ 1, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -234.938  -64.967   -5.102    72.674   311.866
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  410.733      6.482   63.36  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 102.5 on 249 degrees of freedom
```

```
mean(df$agb)
```

```
## [1] 410.7327
```

```
plotrix::std.error(df$agb)
```

```
## [1] 6.482371
```

- a) Beta 0 is equal to 410.7. This is the same as the mean of agb.
- b) The standard error is 6.48 which is the same as the standard error of the sample mean.

Question 2

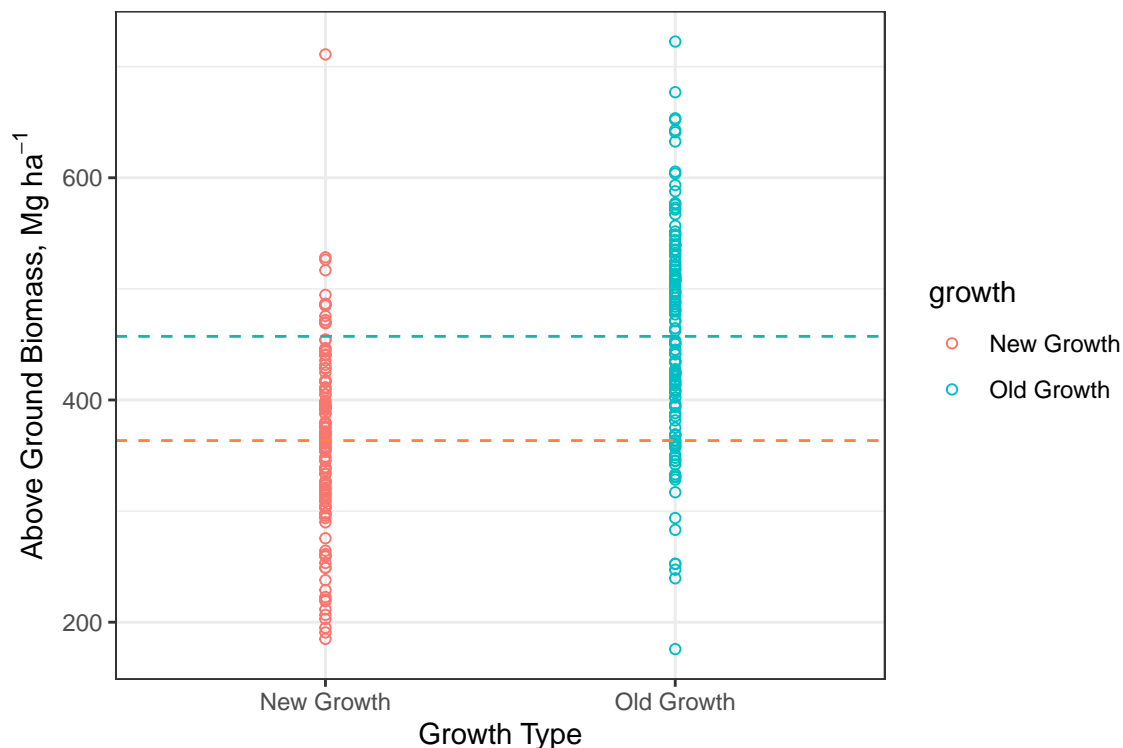
```
lm2 <- lm(formula = agb ~ old_growth, data=df)
summary(lm2)
```

```
##
## Call:
## lm(formula = agb ~ old_growth, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -281.45 -55.02 3.59 54.92 347.54
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 363.469 8.197 44.342 < 2e-16 ***
## old_growth 93.778 11.546 8.122 2.16e-14 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 91.28 on 248 degrees of freedom
## Multiple R-squared: 0.2101, Adjusted R-squared: 0.2069
## F-statistic: 65.97 on 1 and 248 DF, p-value: 2.163e-14
```

- Beta 0 is 363 which is the mean among new old_growth forests
- The difference in the mean between old growth and new growth forests is 93.7
- the value of the estimate for old growth forests is 93.7.
- The value of the estimate of old growth forests is the difference between the average new growth and old growth forest above ground biomass, or the adjustment to the intercept needed for old growth forests.
- the null hypothesis of no difference in agb between old and new growth forests can be rejected with a p value of 2.15e-14
-

```
ggplot(df, aes(x = growth, y = agb, color=growth)) +
  geom_point(shape = 1) +
  geom_hline(yintercept=c(363.469, 363.469+93.778),
             linetype='dashed',
             color=c('coral', 'lightseagreen')) +
  xlab("Growth Type") +
  ylab(expression(paste("Above Ground Biomass, Mg ", ha^-1))) +
  theme_bw()
```



Question 3

```
lm3 <- lm(formula = agb ~ basal_area, data=df)
summary(lm3)
```

```
##
## Call:
## lm(formula = agb ~ basal_area, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -86.975 -36.138   4.273  34.402  87.709
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -113.3818    15.2766  -7.422 1.84e-12 ***
## basal_area     18.8948     0.5422  34.847 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 42.29 on 248 degrees of freedom
## Multiple R-squared:  0.8304, Adjusted R-squared:  0.8297
## F-statistic: 1214 on 1 and 248 DF, p-value: < 2.2e-16
```

```
mean_ba=mean(df$basal_area)
df <- mutate(df, basal_area_centered=basal_area-mean_ba)

lm3.5 <- lm(formula = agb ~ basal_area_centered, data=df)
summary(lm3.5)
```

```
##
## Call:
## lm(formula = agb ~ basal_area_centered, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -86.975 -36.138   4.273  34.402  87.709
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)     410.7327     2.6749  153.55 <2e-16 ***
## basal_area_centered  18.8948     0.5422   34.85 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 42.29 on 248 degrees of freedom
## Multiple R-squared:  0.8304, Adjusted R-squared:  0.8297
## F-statistic: 1214 on 1 and 248 DF, p-value: < 2.2e-16
```

- a) Beta 0 is -113.38. This is what we would expect agb to be when basal area is 0m². This is not particularly meaningful.
- b) Beta 1 is 18.89. This means that when basal area increases by 1m² the average agb is expected to increase by 18.9 Mg/ha
- c) The new value of beta 0 is 410.7, which is the expected abg of samples with average basal area.
- d) Beta 1 is still 18.89 and is still how much abg is expected to increase with 1m² increase in basal area. It

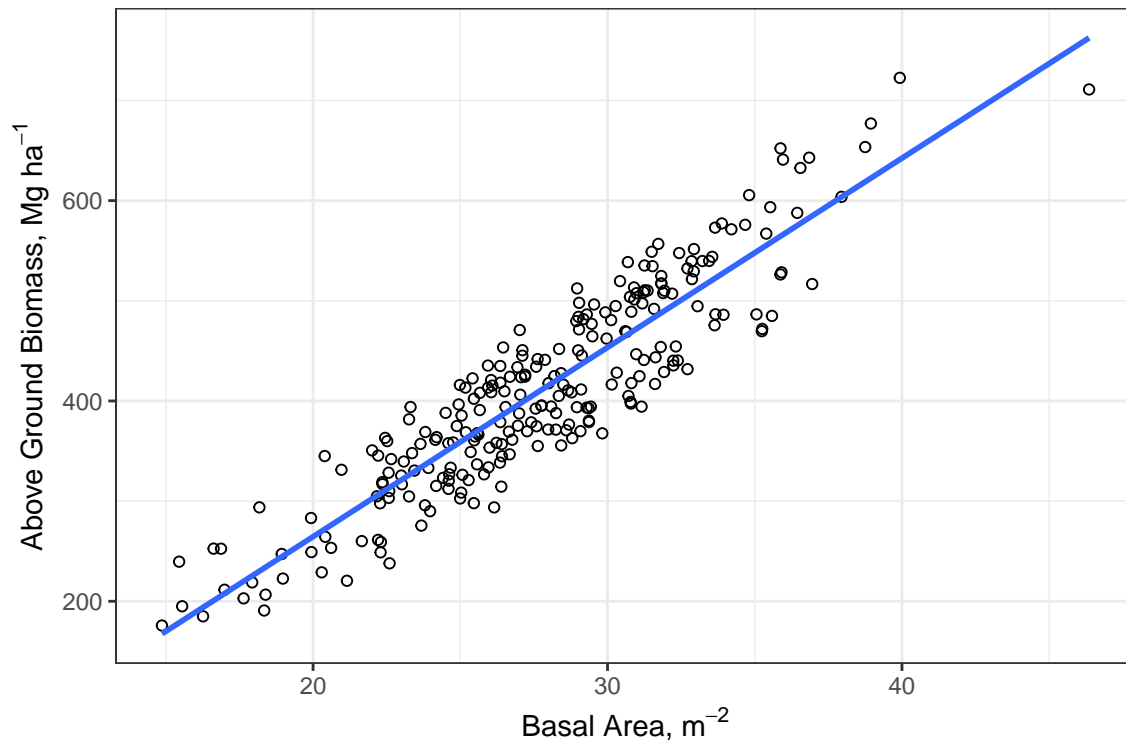
did not change because we only shifted the center of the graph, we did not change the way abg changed with basal area.

e) the null hypothesis of no effect can be rejected with a p value of less than $2e-16$

f)

```
ggplot(df, aes(x = basal_area, y = agb)) +
  geom_point(shape = 1) +
  geom_smooth(method = lm, se=F) +
  xlab(expression(paste("Basal Area, ", m^-2))) +
  ylab(expression(paste("Above Ground Biomass, Mg ", ha^-1))) +
  theme_bw()
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



Question 4

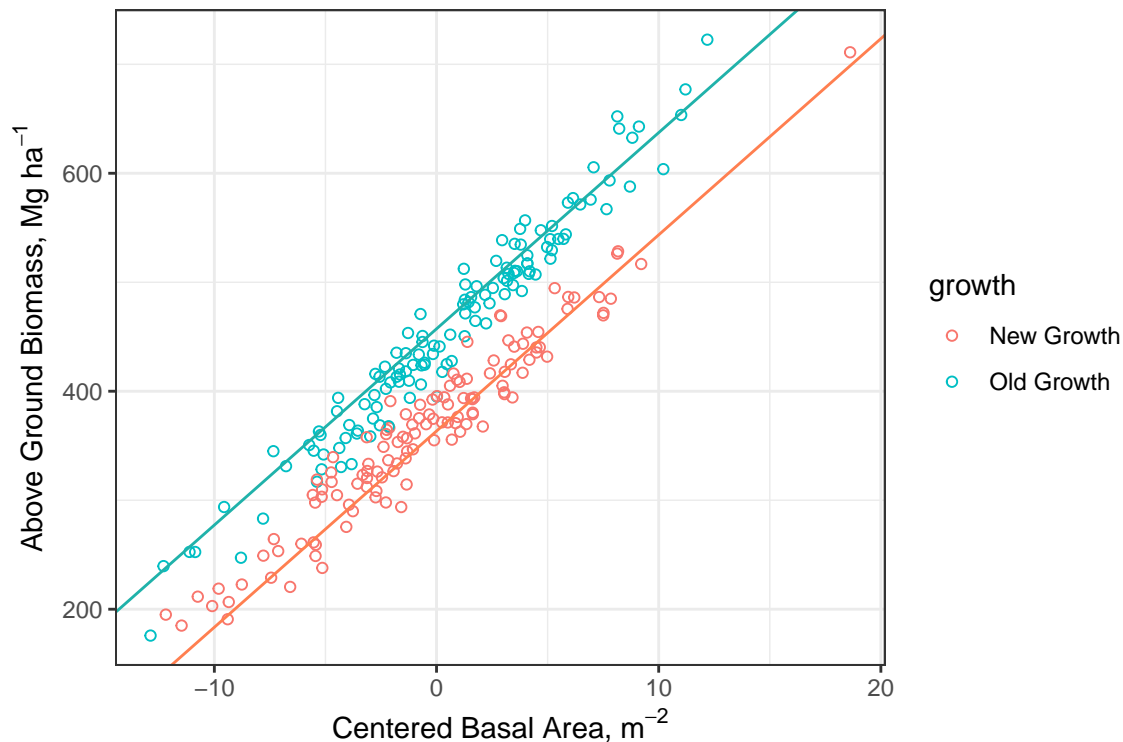
```
lm4 <- lm(formula = agb ~ old_growth + basal_area_centered, data=df)
summary(lm4)
```

```
##
## Call:
## lm(formula = agb ~ old_growth + basal_area_centered, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -52.059  -15.138   -0.136   14.537   59.210
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)          374.399      2.017  185.60   <2e-16 ***
## old_growth           72.090      2.852   25.28   <2e-16 ***
## basal_area_centered  18.003      0.289   62.28   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 22.38 on 247 degrees of freedom
## Multiple R-squared:  0.9527, Adjusted R-squared:  0.9523
## F-statistic: 2488 on 2 and 247 DF,  p-value: < 2.2e-16
```

- Beta 0 is 374 which is the average agb in an new growth forest with average basal area.
- Beta 1 is 72.09 the difference in agb between the average new growth and old growth forest.
- Beta 2 is 18.0. This represents amount abg changes with each additional m2 of basal area across both old and new growth forests.
-

```
ggplot(df, aes(x = basal_area_centered, y = agb, color=growth)) +
  geom_point(shape = 1) +
  geom_abline(intercept=c(363.469, 363.469+93.778),
              slope=c(18.003,18.003),
              color=c('coral', 'lightseagreen'))+
  xlab(expression(paste("Centered Basal Area, ", m^-2))) +
  ylab(expression(paste("Above Ground Biomass, Mg ", ha^-1))) +
  theme_bw()
```



- This plot does look right based on what I was expecting to plot from this model. The results are reasonable and the regression lines fit the data well. Intuitively, something still feels off about both of the sets having the same slope. It just looks a little too tidy.

Question 5

```
df <- mutate(df, old_basal_area_centered=basal_area_centered*old_growth)

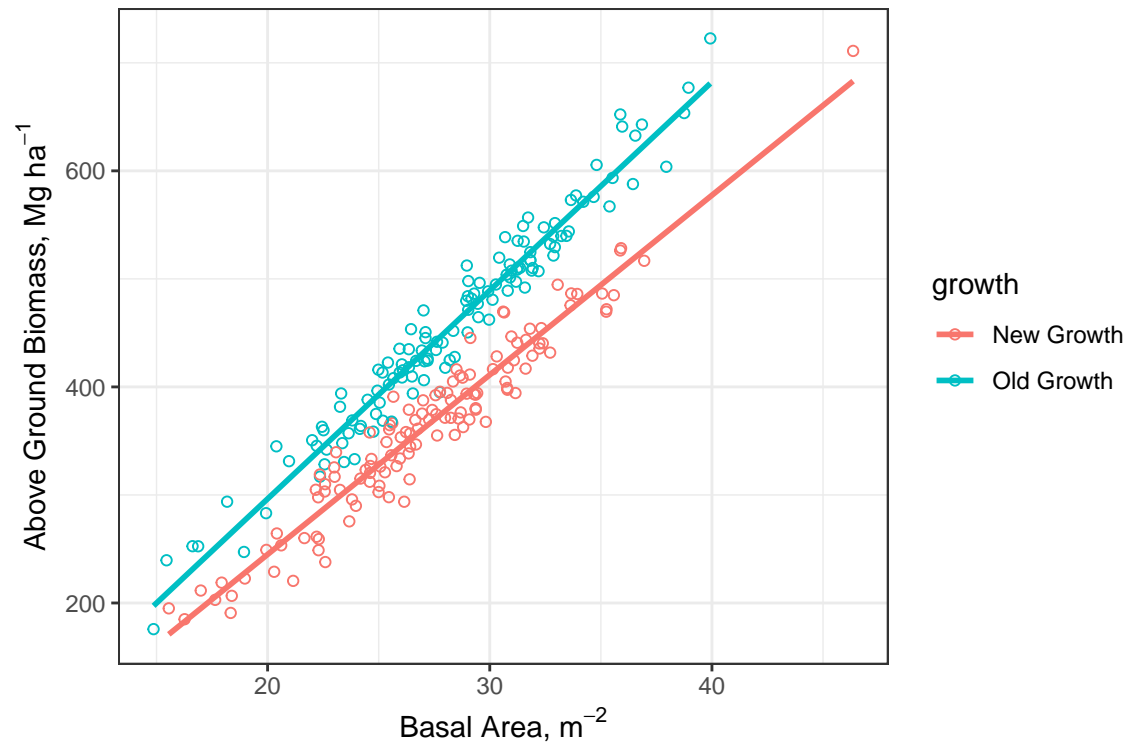
lm5 <- lm(formula = agb ~ old_growth+basal_area_centered+old_basal_area_centered,
          data=df)
summary(lm5)
```

```
##
## Call:
## lm(formula = agb ~ old_growth + basal_area_centered + old_basal_area_centered,
##     data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -53.412 -14.763  -0.841   15.449   51.900
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      373.5625     1.9398  192.578 < 2e-16 ***
## old_growth        72.1542     2.7316   26.415 < 2e-16 ***
## basal_area_centered  16.6245     0.3979   41.781 < 2e-16 ***
## old_basal_area_centered  2.6714     0.5540    4.822 2.49e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 21.43 on 246 degrees of freedom
## Multiple R-squared:  0.9568, Adjusted R-squared:  0.9563
## F-statistic: 1816 on 3 and 246 DF,  p-value: < 2.2e-16
```

- Beta 0 is 373 which is the average abg in new growth forests with average basal area.
- Beta 1 is 72 which is how much higher abg is in old growth forests of average basal area than in new growth forests of average basal area.
- Beta 2 is 16.6 which is the expected change in abg for each 1m² increase in basal area.
- Beta 3 is 2.67 which is the adjustment for the slope to account for the difference in the relationship in abg and basal area among old growth forests.
-

```
ggplot(df, aes(x = basal_area, y = agb, color=growth)) +
  geom_point(shape = 1) +
  geom_smooth(method = lm, se=F) +
  xlab(expression(paste("Basal Area, ", m^-2))) +
  ylab(expression(paste("Above Ground Biomass, Mg ", ha^-1))) +
  theme_bw()
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



f) Again these results look reasonable and like what I would expect. The regression lines seem to fit the data well. It also fits the data slightly better than plot 4 and explains far more of the variation than plot 3.

Question 6

```
df <- mutate(df,
  ca=if_else(state=="CA", 1, 0),
  co=if_else(state=="CO", 1, 0),
  me=if_else(state=="ME", 1, 0),
  nc=if_else(state=="NC", 1, 0))

lm6 <- lm(formula = agb ~ ca+co+me+nc, data=df)
summary(lm6)
```

```
##
## Call:
## lm(formula = agb ~ ca + co + me + nc, data = df)
##
## Residuals:
```

	Min	1Q	Median	3Q	Max
	-237.293	-69.268	-7.507	71.261	286.304

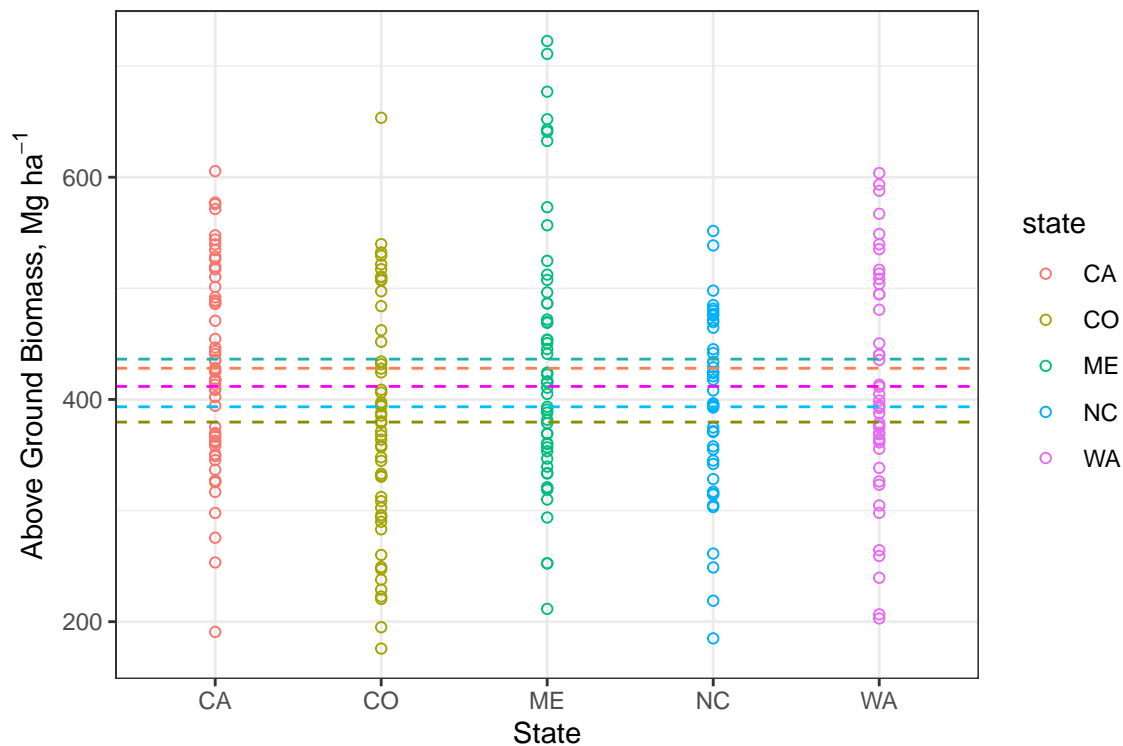
```
##
## Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	411.75	15.06	27.346	<2e-16 ***
ca	16.34	20.14	0.811	0.418

```
## co          -32.11      20.39  -1.575    0.117
## me          24.54      20.47   1.199    0.232
## nc         -18.35      21.81  -0.842    0.401
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 101 on 245 degrees of freedom
## Multiple R-squared:  0.04448,    Adjusted R-squared:  0.02888
## F-statistic: 2.851 on 4 and 245 DF,  p-value: 0.02447
```

- Beta 0 is 411 and this is the expected value of agb in forests in WA.
- Beta 1 is 16 and this is the difference in expected value of agb between forests in CA and WA
- Beta 2 is -32 and this is the difference in expected value of agb between forests in CO and WA
- Beta 3 is 24 and this is the difference in expected value of agb between forests in ME and WA
- Beta 4 is -18 and this is the difference in expected value of agb between forests in NC and WA
-

```
ggplot(df, aes(x = state, y = agb, color=state)) +
  geom_point(shape = 1) +
  geom_hline(
    yintercept=c(411.75, 411.75+16.34,411.75-32.11,411.75+24.54,411.75-18.35),
    linetype='dashed',
    color=c('magenta','coral','yellow4', 'lightseagreen','deepskyblue')) +
  xlab("State") +
  ylab(expression(paste("Above Ground Biomass, Mg ", ha^-1))) +
  theme_bw()
```



- This plot and model show there are some differences between abg between different states but this difference is not particularly great and could also be explained by just sampling differences instead of an inherent difference in the states.