

NEON TOS Coding Lab Solution

K Duffy

8/19/2020

Loading the libraries that we need:

```
library(neonUtilities)
library(geoNEON)
library(sp)
library(dplyr)
library(ggplot2)
```

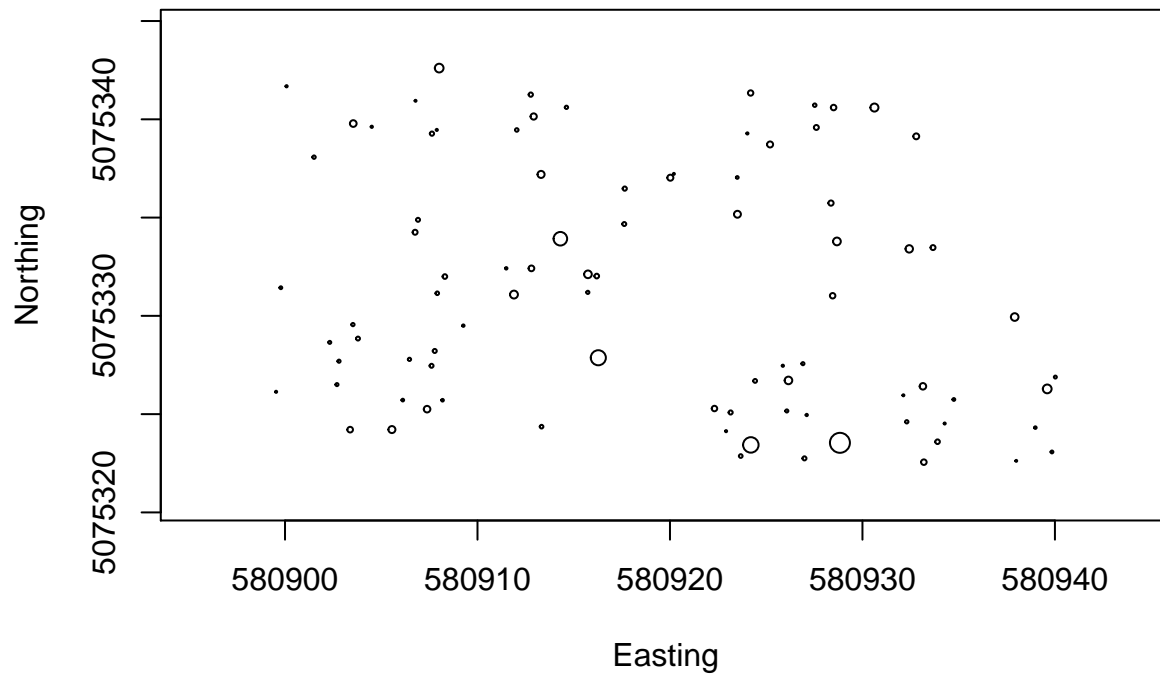
Pulling in the code referenced in the textbook:

```
vegmap <- getLocTOS(veglist$vst_mappingandtagging,
                   "vst_mappingandtagging")
```

```
## |
veg <- merge(veglist$vst_apparentindividual, vegmap,
            by=c("individualID", "namedLocation",
                 "domainID", "siteID", "plotID"))
```

Here's our original plot from the textbook:

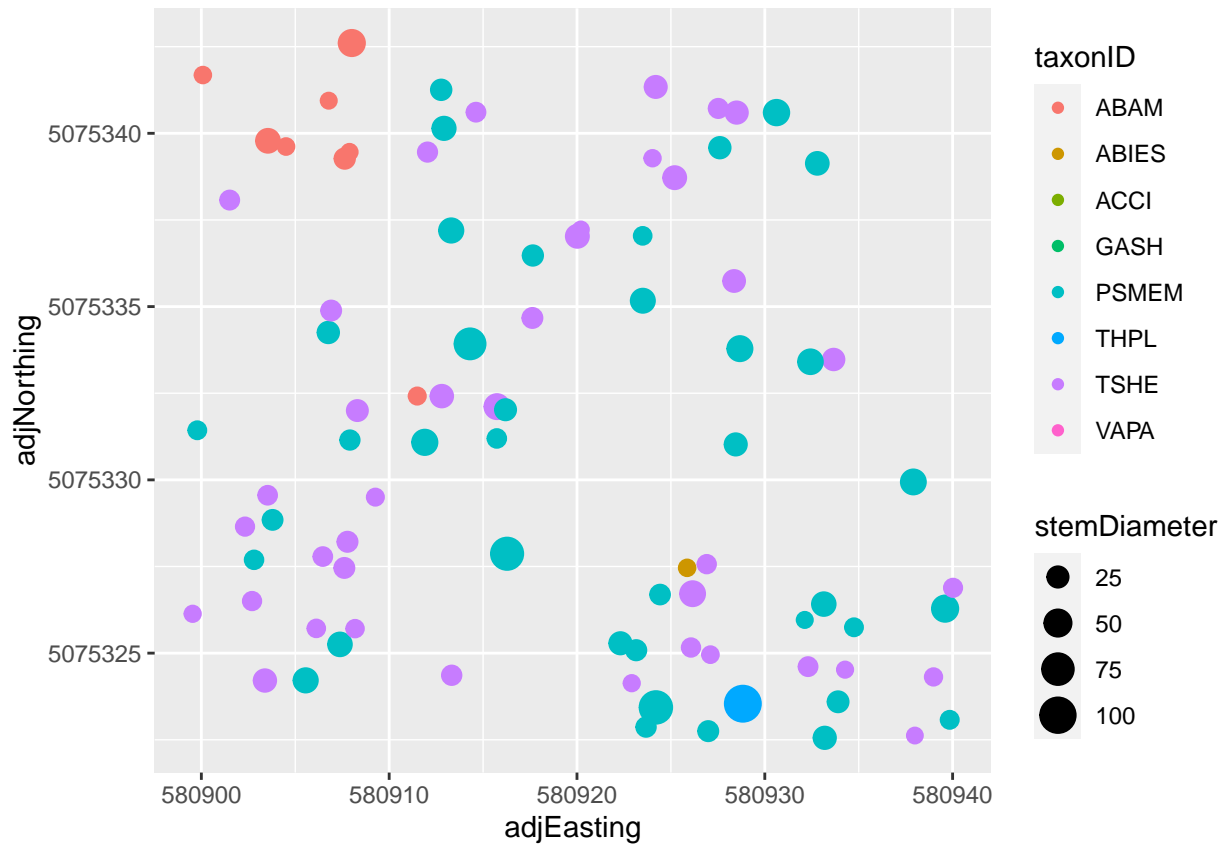
```
symbols(veg$adjEasting[which(veg$plotID=="WREF_075")],
        veg$adjNorthing[which(veg$plotID=="WREF_075")],
        circles=veg$stemDiameter[which(veg$plotID=="WREF_075")]/100/2,
        inches=F, xlab="Easting", ylab="Northing")
```



SOLUTIONS: NEON Part 1

1. Convert the above diameter plot into a ggplot:
2. Set the color of your circles to be a function of each species:

```
ggplot(veg[which(veg$plotID=="WREF_075"),], aes(x=adjEasting, y=adjNorthing, color=taxonID, size=stemDi.  
geom_point()
```



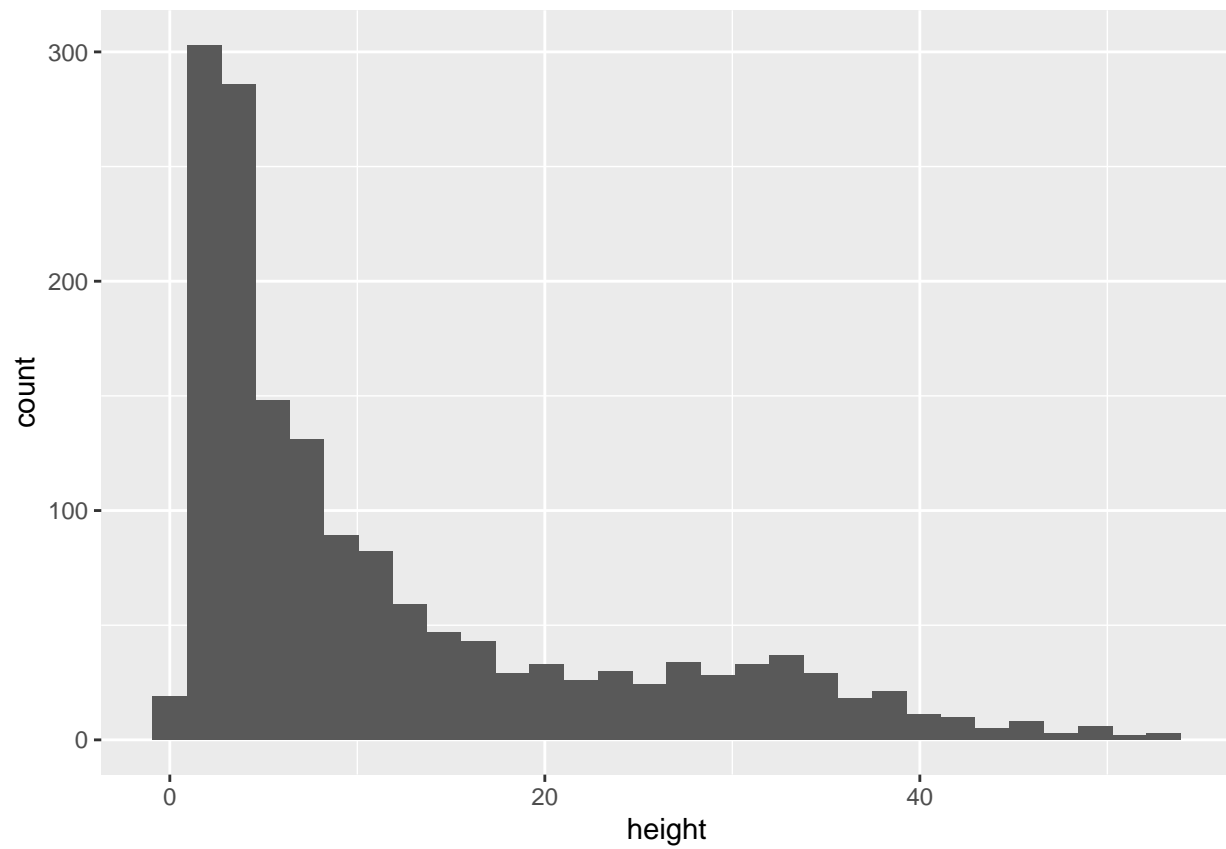
3. Generate a histogram of tree heights for each plot. Color your stacked bar as a function of each species:

First, I'll just make a general histogram:

#I'll remove NA values because they will mess up my model later in this code and give me a bunch of warnings

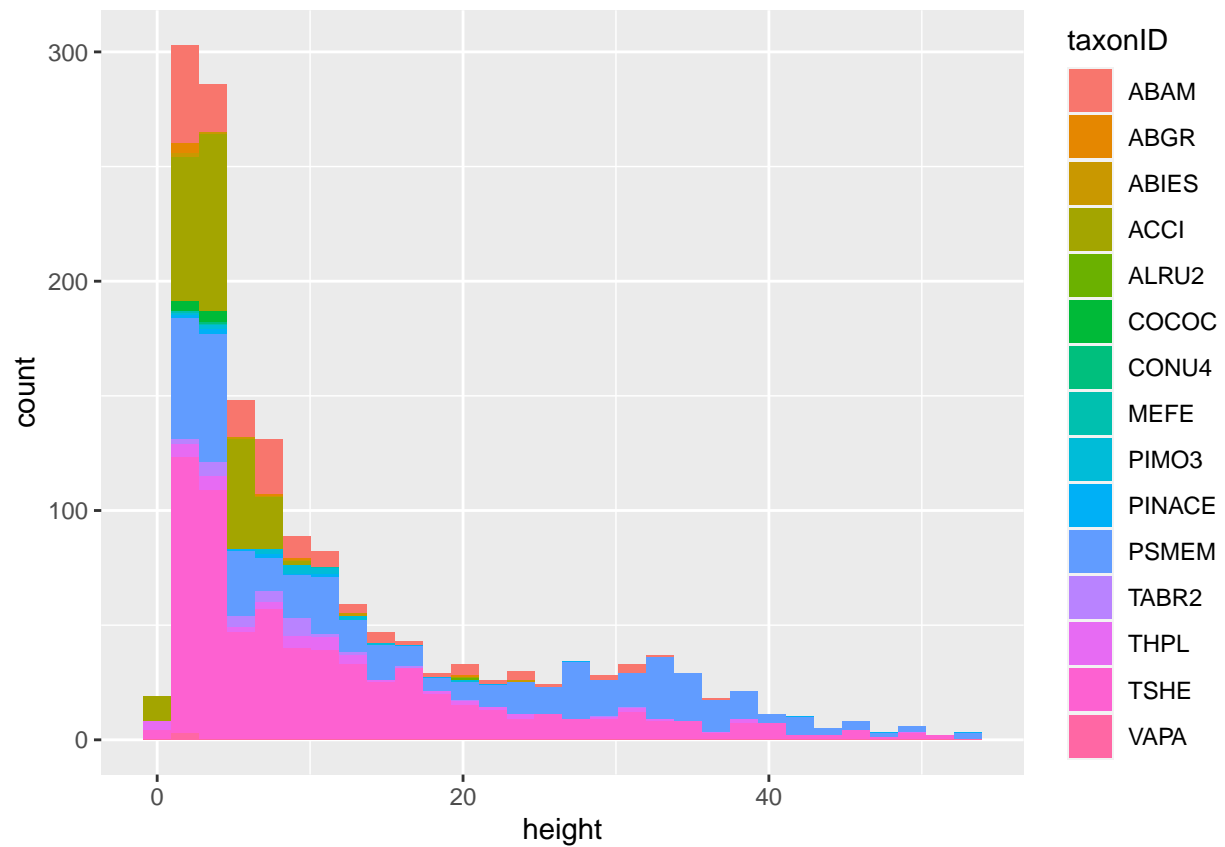
```
veg=veg%>%
  filter(!is.na(stemDiameter))%>%
  filter(!is.na(height))

ggplot(veg, aes(x=height)) +
  geom_histogram()
```



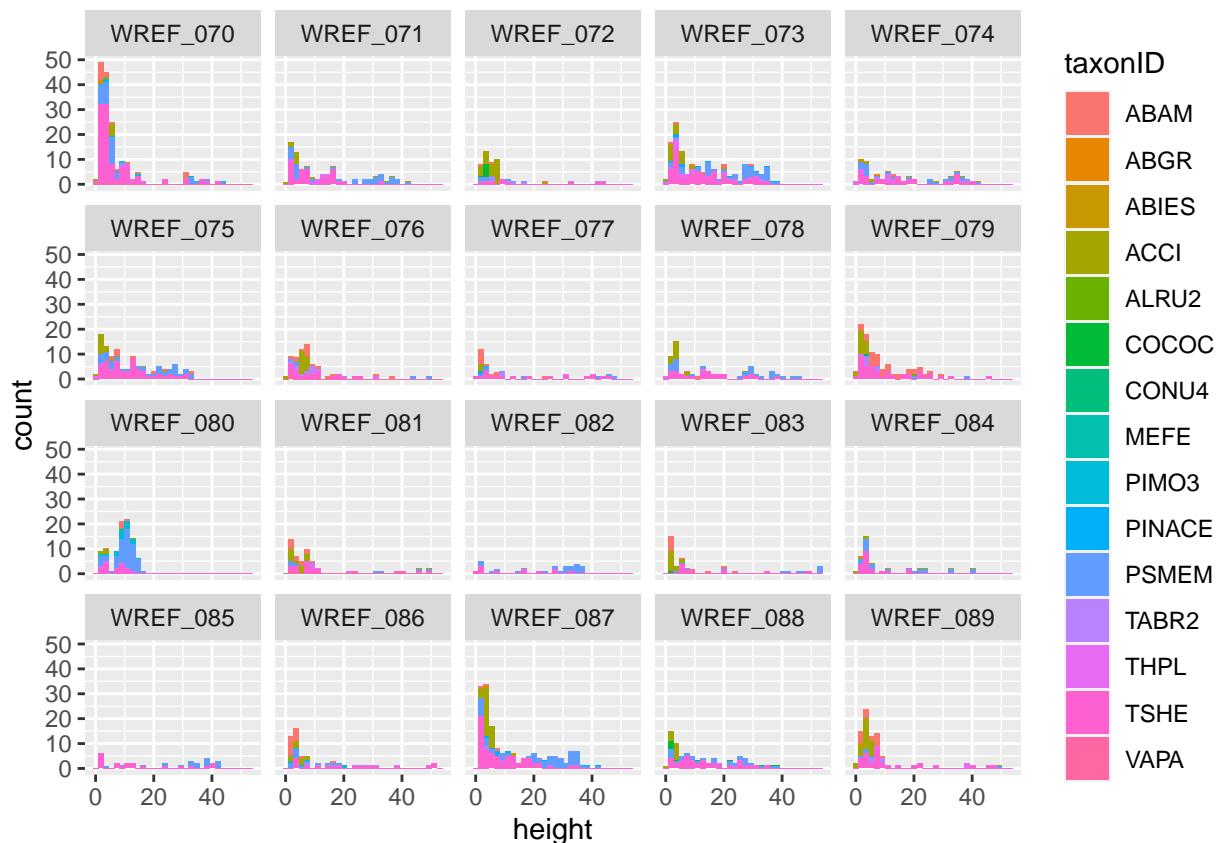
Now I'll add color as a function of species:

```
ggplot(veg, aes(x=height, fill=taxonID)) +  
  geom_histogram()
```



And now I will facet_wrap as a function of plotID:

```
ggplot(veg, aes(x=height, fill=taxonID)) +  
  geom_histogram()+  
  facet_wrap(~ plotID)
```



4. Use `dplyr` to remove dead trees:

First I've looked at the data and metadata. I'm going to figure out all the `unique` values of `plantStatus` at this site:

```
unique(veg$plantStatus)
```

```
## [1] "Dead, broken bole"      "Live"
## [3] "Live, physically damaged" "Standing dead"
## [5] "Live, broken bole"      "Live, disease damaged"
## [7] "Live, other damage"     "Live, insect damaged"
```

Given that check, it looks like I need to filter for "Dead, broken bole" and "Standing dead":

```
vegLIVE=veg%>%
filter(plantStatus != c("Standing dead", "Dead, broken bole"))
```

```
## Warning in plantStatus != c("Standing dead", "Dead, broken bole"): longer object
## length is not a multiple of shorter object length
```

Now a couple of important and quick 'common sense' checks to make sure I really did filter out the dead trees:

First I'll just check how many rows I removed using the base function `nrow` (number of rows)

```
nrow(veg)-nrow(vegLIVE)
```

```
## [1] 145
```

Okay! I removed >100 observations, but I'm getting an error message. Let's just check what all the `plantStatus` values look like now:

```
unique(vegLIVE$plantStatus)
```

```
## [1] "Dead, broken bole"          "Live"
## [3] "Live, physically damaged"    "Live, broken bole"
## [5] "Standing dead"              "Live, disease damaged"
## [7] "Live, other damage"         "Live, insect damaged"
```

Oh no! I still have dead trees! This is *exactly* why we always check.

From here we have a couple of options for filtering. I'll use a `string compare` function from `tidyverse`. Also, I find `dplyr::filter` to be problematic when searching for multiple things at the same time, so I've broken those filters up for readability/ease:

```
library(tidyverse)
vegLIVE=vegLIVE %>%
  filter(!str_detect(plantStatus, 'Standing dead'))%>%
  filter(!str_detect(plantStatus, 'Dead, broken bole'))
```

Let's do another common sense check:

```
unique(vegLIVE$plantStatus)
```

```
## [1] "Live"                  "Live, physically damaged"
## [3] "Live, broken bole"     "Live, disease damaged"
## [5] "Live, other damage"    "Live, insect damaged"
```

No dead trees, success!

To make things a little easier from here on out I'll just plot the top 3 species

```
vegTop3=vegLIVE%>%
  count(vegLIVE$taxonID) %>%
  top_n(3)
```

Selecting by n

```
summary(vegTop3)
```

```
##  vegLIVE$taxonID      n
## Length:3           Min.   :202.0
## Class :character    1st Qu.:244.0
## Mode  :character    Median :286.0
##                               Mean  :356.7
##                               3rd Qu.:434.0
##                               Max.   :582.0
```

Now I'll filter my main data.frame based on this top 3 data.frame:

```
vegLIVE=vegLIVE%>%
  filter(is.element(vegLIVE$taxonID, vegTop3$`vegLIVE$taxonID`))
```

5. Create a simple linear model that uses DBH and height to predict allometries. Print the summary information of your model:

```
mdl=lm(height ~ stemDiameter * taxonID, data=vegLIVE)
print(mdl)
```

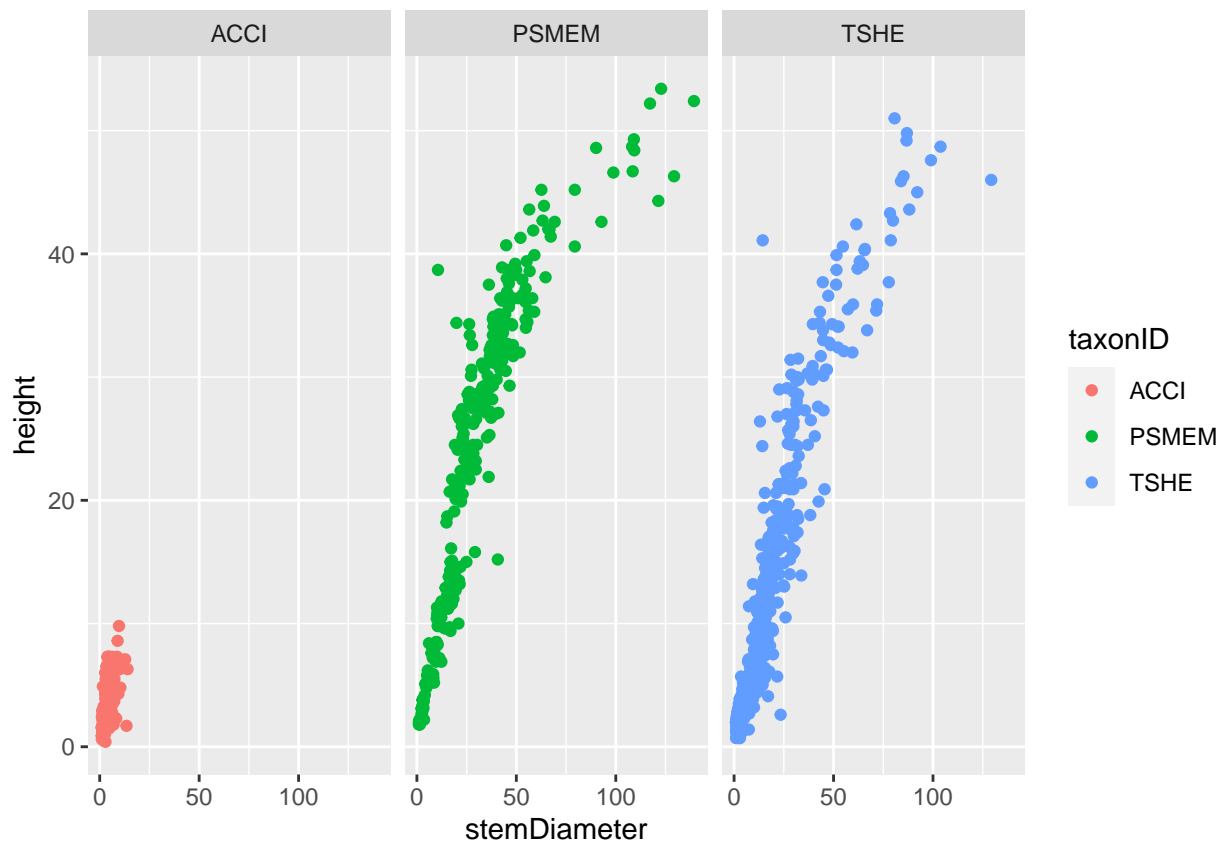
```
##
## Call:
## lm(formula = height ~ stemDiameter * taxonID, data = vegLIVE)
##
```

```
## Coefficients:
##              (Intercept)              stemDiameter
##              1.91986              0.44364
##              taxonIDPSMEM              taxonIDTSHE
##              7.68473              0.32365
## stemDiameter:taxonIDPSMEM stemDiameter:taxonIDTSHE
##              0.01772              0.12846
```

Now I'll tack on my model predictions to my data.frame:

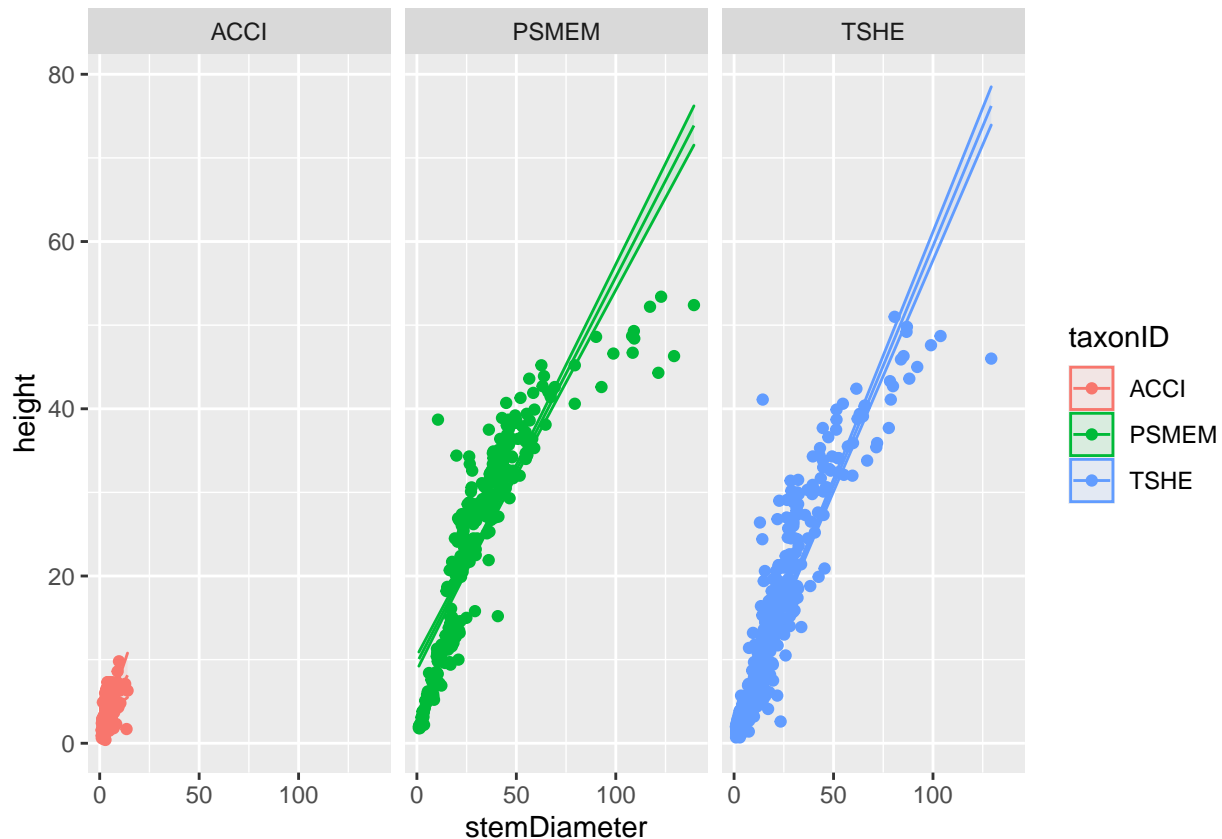
Let's take a look. First let's plot our NEON data:

```
ggplot(data=vegLIVE, aes(x=stemDiameter, y=height, color=taxonID)) +
  geom_point() +
  facet_wrap(~ taxonID)
```



Now let's plot our model:

```
ggplot(data=vegLIVE, aes(x=stemDiameter, y=height, color=taxonID)) +
  geom_point() +
  geom_line( aes(y=fit) ) +
  geom_ribbon( aes( ymin=lwr, ymax=upr, fill=taxonID), alpha=.1 ) +
  facet_wrap(~ taxonID) # alpha is the ribbon transparency
```

What do you think about our simple linear model? To me it looks like we're failing to constrain the fact that trees have a height limit. I guess gravity isn't just a 'theory'. In this case I would certainly look at non-linear methods such as splines or a GAM, but we'll apply those methods later in this course.

7. Answer the following questions:

- How many unique species are present at WREF?

```
print(length(unique(veg$taxonID)))
```

```
## [1] 15
```

- What are the top_5 trees based on height?

```
vegTop5_height=veg%>%
  slice_max(height, n=5)
head(vegTop5_height)
```

```
##      individualID      namedLocation domainID siteID  plotID
## 1 NEON.PLA.D16.WREF.04807 WREF_083.basePlot.vst      D16   WREF WREF_083
## 2 NEON.PLA.D16.WREF.04803 WREF_083.basePlot.vst      D16   WREF WREF_083
## 3 NEON.PLA.D16.WREF.04934 WREF_083.basePlot.vst      D16   WREF WREF_083
## 4 NEON.PLA.D16.WREF.04159 WREF_086.basePlot.vst      D16   WREF WREF_086
## 5 NEON.PLA.D16.WREF.04148 WREF_086.basePlot.vst      D16   WREF WREF_086
##      uid.x      date.x      eventID.x subplotID.x
## 1 04151931-5979-4e33-8c10-d1b6b6ddb745 2017-10-18 vst_WREF_2017      <NA>
## 2 3da7c3bc-aead-4c2a-9838-7eb1c83ddb745 2017-10-18 vst_WREF_2017      <NA>
## 3 b79e1d0d-4eea-43c3-8a2a-3f5cf12e3601 2017-10-18 vst_WREF_2017      <NA>
## 4 fa7f9079-5489-4453-9ad3-be96aea28c30 2017-10-20 vst_WREF_2017      <NA>
## 5 cb48c34a-c4bd-4bb8-8a41-f4a04b240f32 2017-10-20 vst_WREF_2017      <NA>
```

##	tempStemID	tagStatus	growthForm	plantStatus	stemDiameter
## 1	<NA>	ok	single bole tree	Live	122.8
## 2	<NA>	ok	single bole tree	Live	139.3
## 3	<NA>	ok	single bole tree	Live	117.2
## 4	<NA>	ok	single bole tree	Live	80.7
## 5	<NA>	ok	single bole tree	Standing dead	83.0
##	measurementHeight	height	baseCrownHeight	breakHeight	breakDiameter
## 1	130	53.4	NA	NA	NA
## 2	130	52.4	NA	NA	NA
## 3	130	52.2	NA	NA	NA
## 4	130	51.0	NA	NA	NA
## 5	180	50.8	NA	NA	NA
##	maxCrownDiameter	ninetyCrownDiameter	canopyPosition	shape	basalStemDiameter
## 1	NA	NA	<NA>	<NA>	NA
## 2	NA	NA	<NA>	<NA>	NA
## 3	NA	NA	<NA>	<NA>	NA
## 4	NA	NA	<NA>	<NA>	NA
## 5	NA	NA	<NA>	<NA>	NA
##	basalStemDiameter	msrmntHeight	maxBaseCrownDiameter	ninetyBaseCrownDiameter	
## 1		NA	NA	NA	
## 2		NA	NA	NA	
## 3		NA	NA	NA	
## 4		NA	NA	NA	
## 5		NA	NA	NA	
##	dendrometerInstallationDate	initialGapMeasurementDate	initialBandStemDiameter		
## 1		NA	NA		NA
## 2		NA	NA		NA
## 3		NA	NA		NA
## 4		NA	NA		NA
## 5		NA	NA		NA
##	initialDendrometerGap	dendrometerHeight	dendrometerGap	dendrometerCondition	
## 1		NA	NA	NA	<NA>
## 2		NA	NA	NA	<NA>
## 3		NA	NA	NA	<NA>
## 4		NA	NA	NA	<NA>
## 5		NA	NA	NA	<NA>
##	bandStemDiameter	remarks.x	recordedBy.x		
## 1	NA	<NA>	dwyer@battelleecology.org		
## 2	NA	<NA>	dwyer@battelleecology.org		
## 3	NA	<NA>	jhausknecht@battelleecology.org		
## 4	NA	Dwarf mistletoe	ccordell@battelleecology.org		
## 5	NA	Dbh blocked by fallen tree	ccordell@battelleecology.org		
##	measuredBy.x	dataQF.x	publicationDate.x		
## 1	vschmitt@battelleecology.org	legacyData	20200817T130556Z		
## 2	vschmitt@battelleecology.org	legacyData	20200817T130556Z		
## 3	Kchilders@battelleecology.org	legacyData	20200817T130556Z		
## 4	Eolsen@battelleecology.org	legacyData	20200817T130556Z		
## 5	Eolsen@battelleecology.org	legacyData	20200817T130556Z		
##	uid.y	date.y	eventID.y	subplotID.y	
## 1	4529d5b1-6a8b-4e94-b471-4cc8ba72c067	2017-10-18	vst_WREF_2017		21
## 2	8e9eb50f-44bf-4b99-833a-09cf9c4022ff	2017-10-18	vst_WREF_2017		21
## 3	f6db7136-9b6e-49d6-8271-93961979574f	2017-10-18	vst_WREF_2017		23
## 4	86ea41d6-b74e-43fe-a675-11d269235956	2017-10-23	vst_WREF_2017		21
## 5	efc3dbad-9f54-4052-8b79-abd5f0dec396	2017-10-23	vst_WREF_2017		21

```

##      nestedSubplotID pointID stemDistance stemAzimuth recordType
## 1              4      41          4.0        248.5        <NA>
## 2             <NA>      39          2.3        102.0        <NA>
## 3              2      25          7.3        313.3        <NA>
## 4              1      23         11.5        299.6        <NA>
## 5              4      41         10.5        223.1        <NA>
##      supportingStemIndividualID previouslyTaggedAs samplingProtocolVersion taxonID
## 1                      <NA>          <NA>      NEON.DOC.000987vG   PSMEM
## 2                      <NA>          <NA>      NEON.DOC.000987vG   PSMEM
## 3                      <NA>          <NA>      NEON.DOC.000987vG   PSMEM
## 4                      <NA>          <NA>      NEON.DOC.000987vG   TSHE
## 5                      <NA>          <NA>      NEON.DOC.000987vG   TSHE
##                                     scientificName taxonRank
## 1 Pseudotsuga menziesii (Mirb.) Franco var. menziesii   variety
## 2 Pseudotsuga menziesii (Mirb.) Franco var. menziesii   variety
## 3 Pseudotsuga menziesii (Mirb.) Franco var. menziesii   variety
## 4              Tsuga heterophylla (Raf.) Sarg.   species
## 5              Tsuga heterophylla (Raf.) Sarg.   species
##      identificationReferences morphospeciesID morphospeciesIDRemarks
## 1                      <NA>          <NA>          <NA>
## 2                      <NA>          <NA>          <NA>
## 3                      <NA>          <NA>          <NA>
## 4                      <NA>          <NA>          <NA>
## 5                      <NA>          <NA>          <NA>
##      identificationQualifier remarks.y          measuredBy.y
## 1                      <NA>      <NA>      Eolsen@battelleecology.org
## 2                      <NA>      <NA>      Eolsen@battelleecology.org
## 3                      <NA>      <NA>      Kchilders@battelleecology.org
## 4                      <NA>      <NA>      Eolsen@battelleecology.org
## 5                      <NA>      <NA>      Eolsen@battelleecology.org
##                                     recordedBy.y dataQF.y publicationDate.y utmZone
## 1      vschmitt@battelleecology.org      <NA> 20200817T124831Z      10N
## 2      vschmitt@battelleecology.org      <NA> 20200817T124831Z      10N
## 3      jhausknecht@battelleecology.org      <NA> 20200817T124831Z      10N
## 4      Kchilders@battelleecology.org      <NA> 20200817T124831Z      10N
## 5      Kchilders@battelleecology.org      <NA> 20200817T124831Z      10N
##      adjNorthing adjEasting adjCoordinateUncertainty adjDecimallLatitude
## 1      5074889      580560.2          0.83          45.82286
## 2      5074888      580545.8          0.82          45.82285
## 3      5074876      580579.4          0.82          45.82274
## 4      5074818      581000.5          0.83          45.82217
## 5      5074823      581003.9          0.83          45.82221
##      adjDecimallLongitude adjElevation adjElevationUncertainty
## 1      -121.9629          396.87          1.35
## 2      -121.9631          399.79          1.26
## 3      -121.9627          401.64          1.32
## 4      -121.9573          378.39          1.35
## 5      -121.9572          374.47          1.26
vegTop5_height%>%
  select(plotID, plantStatus, taxonID, scientificName, adjNorthing, adjEasting, height)

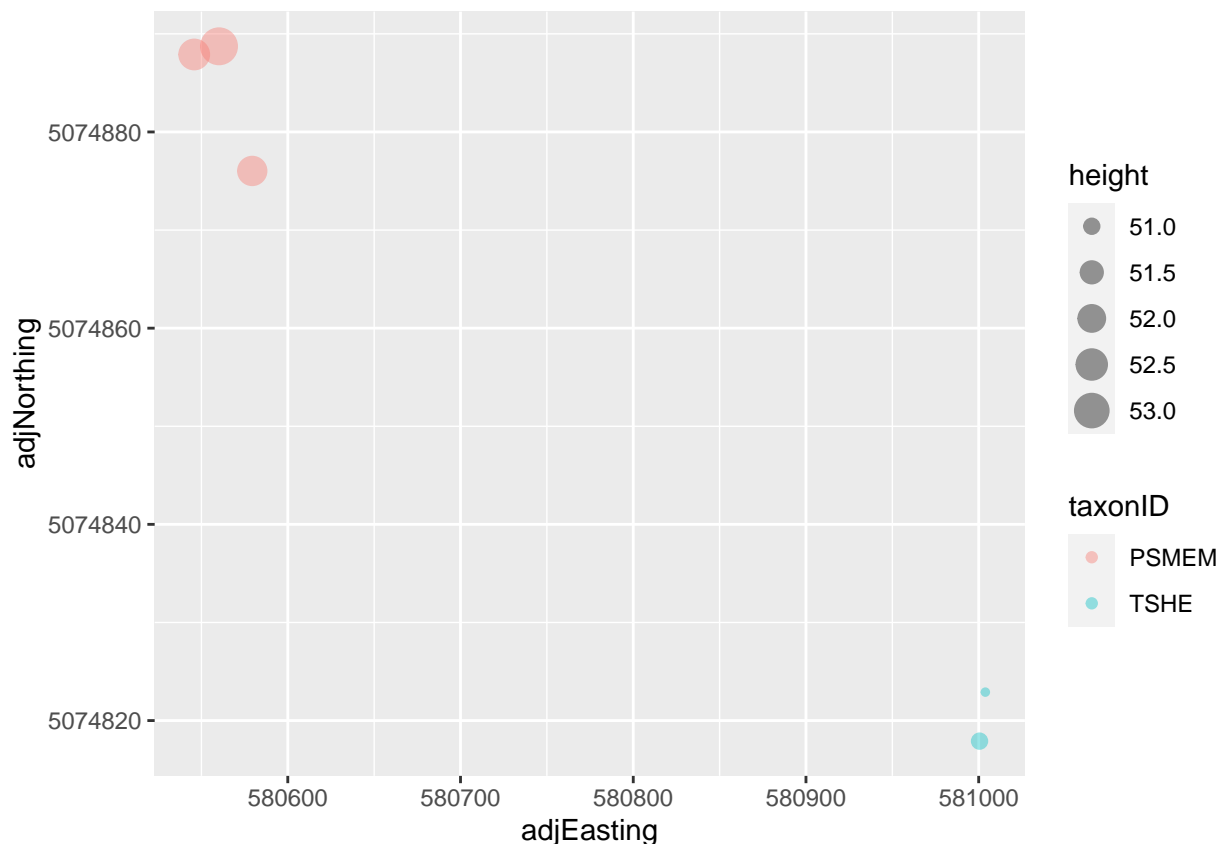
##      plotID      plantStatus taxonID
## 1 WREF_083      Live      PSMEM
## 2 WREF_083      Live      PSMEM

```

```
## 3 WREF_083          Live   PSMEM
## 4 WREF_086          Live   TSHE
## 5 WREF_086 Standing dead   TSHE
##
##               scientificName adjNorthing adjEasting
## 1 Pseudotsuga menziesii (Mirb.) Franco var. menziesii 5074889 580560.2
## 2 Pseudotsuga menziesii (Mirb.) Franco var. menziesii 5074888 580545.8
## 3 Pseudotsuga menziesii (Mirb.) Franco var. menziesii 5074876 580579.4
## 4                Tsuga heterophylla (Raf.) Sarg.    5074818 581000.5
## 5                Tsuga heterophylla (Raf.) Sarg.    5074823 581003.9
##   height
## 1   53.4
## 2   52.4
## 3   52.2
## 4   51.0
## 5   50.8
```

Cool. My native state tree, the Douglas Fir, dominates the height record (Go Oregon!). I wonder if these trees are all together in an old growth stand, or if they're distributed throughout the area?

```
ggplot(vegTop5_height, aes(x=adjEasting, y=adjNorthing, color=taxonID, size=height)) +
  geom_point(alpha=.4 )
```



It looks to me like we have a small old growth patch of Douglas Fir. That would be super fun to study.

- Diameter?

```
vegTop5_diameter=veg%>%
  slice_max(stemDiameter, n=5)
head(vegTop5_diameter)
```

##	individualID	namedLocation	domainID	siteID	plotID
## 1	NEON.PLA.D16.WREF.04360	WREF_086.basePlot.vst	D16	WREF	WREF_086
## 2	NEON.PLA.D16.WREF.04803	WREF_083.basePlot.vst	D16	WREF	WREF_083
## 3	NEON.PLA.D16.WREF.04747	WREF_086.basePlot.vst	D16	WREF	WREF_086
## 4	NEON.PLA.D16.WREF.02139	WREF_079.basePlot.vst	D16	WREF	WREF_079
## 5	NEON.PLA.D16.WREF.04509	WREF_077.basePlot.vst	D16	WREF	WREF_077
##	uid.x	date.x	eventID.x	subplotID.x	
## 1	5bc4101d-24c8-491e-8e97-4aa456b235ee	2017-10-20	vst_WREF_2017		<NA>
## 2	3da7c3bc-aead-4c2a-9838-7eb1c83ddb4	2017-10-18	vst_WREF_2017		<NA>
## 3	da1661cf-d9b8-41b6-a887-8afb76be159e	2017-10-20	vst_WREF_2017		<NA>
## 4	fcf2924f-94c3-446f-8b0d-07b45fb52d55	2017-10-26	vst_WREF_2017		<NA>
## 5	72704d5e-e0cd-4fab-b683-d6eb26a52cc6	2017-10-18	vst_WREF_2017		<NA>
##	tempStemID	tagStatus	growthForm	plantStatus	stemDiameter
## 1	<NA>	ok	single bole tree	Dead, broken bole	144.2
## 2	<NA>	ok	single bole tree	Live	139.3
## 3	<NA>	ok	single bole tree	Dead, broken bole	131.5
## 4	<NA>	ok	single bole tree	Live, disease damaged	129.3
## 5	<NA>	ok	single bole tree	Live	129.3
##	measurementHeight	height	baseCrownHeight	breakHeight	breakDiameter
## 1	130	19.1	NA	NA	NA
## 2	130	52.4	NA	NA	NA
## 3	130	3.9	NA	NA	NA
## 4	130	46.0	NA	NA	NA
## 5	130	46.3	NA	NA	NA
##	maxCrownDiameter	ninetyCrownDiameter	canopyPosition	shape	basalStemDiameter
## 1	NA	NA	<NA>	<NA>	NA
## 2	NA	NA	<NA>	<NA>	NA
## 3	NA	NA	<NA>	<NA>	NA
## 4	NA	NA	<NA>	<NA>	NA
## 5	NA	NA	<NA>	<NA>	NA
##	basalStemDiameterMsrmtHeight	maxBaseCrownDiameter	ninetyBaseCrownDiameter		
## 1	NA	NA	NA		NA
## 2	NA	NA	NA		NA
## 3	NA	NA	NA		NA
## 4	NA	NA	NA		NA
## 5	NA	NA	NA		NA
##	dendrometerInstallationDate	initialGapMeasurementDate	initialBandStemDiameter		
## 1	NA	NA	NA		NA
## 2	NA	NA	NA		NA
## 3	NA	NA	NA		NA
## 4	NA	NA	NA		NA
## 5	NA	NA	NA		NA
##	initialDendrometerGap	dendrometerHeight	dendrometerGap	dendrometerCondition	
## 1	NA	NA	NA	<NA>	
## 2	NA	NA	NA	<NA>	
## 3	NA	NA	NA	<NA>	
## 4	NA	NA	NA	<NA>	
## 5	NA	NA	NA	<NA>	
##	bandStemDiameter	remarks.x			
## 1	NA	Broken main bole			
## 2	NA	<NA>			
## 3	NA	<NA>			
## 4	NA	Heavily infested with Dwarf mistletoe			
## 5	NA	<NA>			

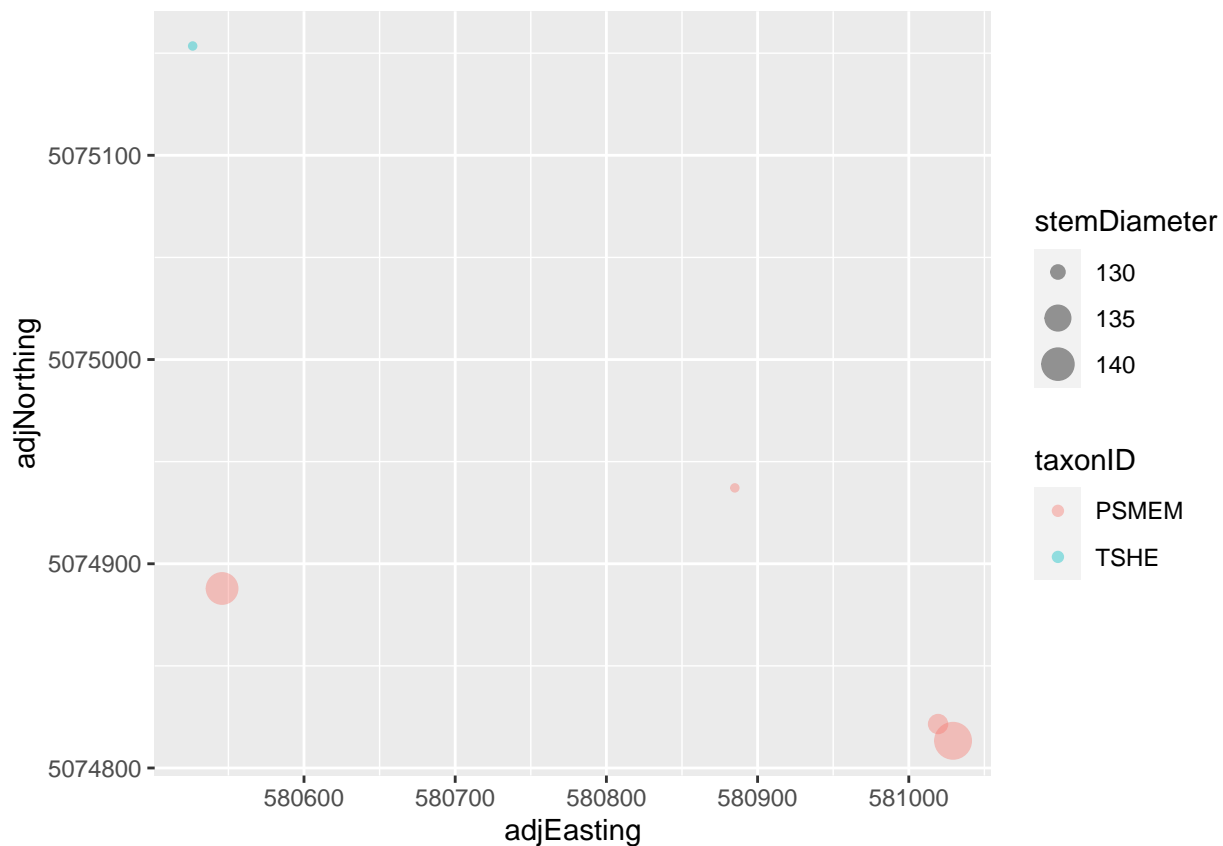
```

##               recordedBy.x               measuredBy.x      dataQF.x
## 1 jhausknecht@battelleecology.org Eolsen@battelleecology.org legacyData
## 2 dweller@battelleecology.org vschmitt@battelleecology.org legacyData
## 3 jhausknecht@battelleecology.org Eolsen@battelleecology.org legacyData
## 4 Kchilders@battelleecology.org kwellls@battelleecology.org legacyData
## 5 dweller@battelleecology.org Eolsen@battelleecology.org legacyData
##   publicationDate.x               uid.y      date.y
## 1 20200817T130556Z 68f2c91b-0bfc-4497-8270-ebb6e8d7db26 2017-10-20
## 2 20200817T130556Z 8e9eb50f-44bf-4b99-833a-09cf9c4022ff 2017-10-18
## 3 20200817T130556Z ea25b600-0492-4237-908b-c32d2f9c05c8 2017-10-23
## 4 20200817T130556Z 460a43cf-3ddb-449a-b540-bb98cff80fd1 2017-10-26
## 5 20200817T130556Z 41992938-8784-4d23-8b7d-ac701db9a358 2017-10-19
##   eventID.y subplotID.y nestedSubplotID pointID stemDistance stemAzimuth
## 1 vst_WREF_2017      23          2      41          25.1      133.4
## 2 vst_WREF_2017      21          <NA>     39          2.3      102.0
## 3 vst_WREF_2017      23          <NA>     41          12.3     137.3
## 4 vst_WREF_2017      21          4      41          7.7      197.5
## 5 vst_WREF_2017      39          <NA>     41          14.7     343.0
##   recordType supportingStemIndividualID previouslyTaggedAs
## 1      <NA>                <NA>                <NA>
## 2      <NA>                <NA>                <NA>
## 3      <NA>                <NA>                <NA>
## 4      <NA>                <NA>                <NA>
## 5      <NA>                <NA>                <NA>
##   samplingProtocolVersion taxonID
## 1 NEON.DOC.000987vG  PSMEM
## 2 NEON.DOC.000987vG  PSMEM
## 3 NEON.DOC.000987vG  PSMEM
## 4 NEON.DOC.000987vG  TSHE
## 5 NEON.DOC.000987vG  PSMEM
##                                     scientificName taxonRank
## 1 Pseudotsuga menziesii (Mirb.) Franco var. menziesii  variety
## 2 Pseudotsuga menziesii (Mirb.) Franco var. menziesii  variety
## 3 Pseudotsuga menziesii (Mirb.) Franco var. menziesii  variety
## 4 Tsuga heterophylla (Raf.) Sarg.  species
## 5 Pseudotsuga menziesii (Mirb.) Franco var. menziesii  variety
##   identificationReferences morphospeciesID morphospeciesIDRemarks
## 1      <NA>                <NA>                <NA>
## 2      <NA>                <NA>                <NA>
## 3      <NA>                <NA>                <NA>
## 4      <NA>                <NA>                <NA>
## 5      <NA>                <NA>                <NA>
##   identificationQualifier remarks.y               measuredBy.y
## 1      <NA>      <NA>      Eolsen@battelleecology.org
## 2      <NA>      <NA>      Eolsen@battelleecology.org
## 3      <NA>      <NA>      ccordell@battelleecology.org
## 4      <NA>      <NA>      kwellls@battelleecology.org
## 5      <NA>      <NA>      Eolsen@battelleecology.org
##   recordedBy.y dataQF.y publicationDate.y utmZone
## 1 jhausknecht@battelleecology.org <NA> 20200817T124831Z 10N
## 2 vschmitt@battelleecology.org <NA> 20200817T124831Z 10N
## 3 dweller@battelleecology.org <NA> 20200817T124831Z 10N
## 4 Kchilders@battelleecology.org <NA> 20200817T124831Z 10N
## 5 dweller@battelleecology.org <NA> 20200817T124831Z 10N

```

```
##   adjNorthing adjEasting adjCoordinateUncertainty adjDecimalLatitude
## 1    5074813   581029.3             0.83             45.82212
## 2    5074888   580545.8             0.82             45.82285
## 3    5074822   581019.4             0.83             45.82220
## 4    5075154   580526.4             0.77             45.82524
## 5    5074937   580884.9             0.80             45.82325
##   adjDecimalLongitude adjElevation adjElevationUncertainty
## 1             -121.9569         374.47             1.26
## 2             -121.9631         399.79             1.26
## 3             -121.9570         374.47             1.26
## 4             -121.9633         390.5             1.16
## 5             -121.9587         379.83             1.28
```

```
ggplot(vegTop5_diameter, aes(x=adjEasting, y=adjNorthing, color=taxonID, size=stemDiameter)) +
  geom_point(alpha=.4 )
```



Interesting. Also mostly Douglas Firs, but not the ones who are the tallest. I bet that's why our linear model failed.

- What proportion of sampled trees are dead?

```
length(veg)/length(vegLIVE)
```

```
## [1] 0.96
```

Or ~4%