Data cleaning and exploration

Jody Daniel

2020-11-03

Contents

Traits and Plant Growth Rates

We have two datasets: 1) basal area each year from 2006 to 2011 and 2) plant and environmental traits for each tree. Below, I aim to examine the data - assess data types, check for missingness, normality, outliers. Depending on the degree of missingness, I will impute the data as we would like to keep as many observations as possible.

Table of Contents

- Initial Assessment
- Imputation
- Estimate Growth Rate
- Conclusion

```
library(corrplot)
library(fastDummies)
library(RColorBrewer)
library(factoextra)
library(ggplot2)
require(ggrepel)
library(knitr)
library(kableExtra)
library(tidyverse)
library(dplyr)
library(here)
library(skimr)
library(reshape2)
```

```
library(tidymodels)
library(qdapTools)
library(rsample)
library(corrr)
library(broom)
library(vegan)
library(extrafont)
library(viridis)
library(car)
library(mice)
library(sjmisc)
library(skimr)
library(RVAideMemoire)
source(here("scripts/archive/1. functions.R"))
theme_set(theme_special())
```

Extracting Principle Components for Environmental Traits

```
rgr_raw <- read.csv(here("data/RGR.csv"))
msh_raw <- read.csv(here("data/MSH.csv"))
# what do these data look like?
kable(skim(rgr_raw),"latex", booktabs = T) %>%
  kable_styling(latex_options="scale_down")
```

skim_type	skim_variable	n_missing	complete_rate	character.min	character.max	character.empty	character.n_unique	character.whitespace	numeric.mean	numeric.sd	numeric.p0	numeric.p25	numeric.p50	numeric.p75	numeric.p100	numeric.hist
character	SampleID	0	1.0000000	6	10	0	395	0	NA	NA	NA	NA	NA	NA	NA	NA
numeric	BA.0.2006	9	0.9772152	NA	NA	NA	NA	NA	6.843472	6.296880	0.0000000	2.520196	5.154124	9.99809	50.22676	<U+2587><U+2582><U+2581><U+2581><U+2581>
numeric	BA.0.2007	10	0.9746835	NA	NA	NA	NA	NA	7.784849	6.662033	0.0000000	2.997015	5.926886	11.34191	51.51465	<U+2587><U+2583><U+2581><U+2581><U+2581>
numeric	BA.0.2008	10	0.9746835	NA	NA	NA	NA	NA	8.872749	7.052922	0.0302196	3.730206	7.189076	12.68398	52.24834	<U+2587><U+2583><U+2581><U+2581><U+2581>
numeric	BA.0.2009	10	0.9746835	NA	NA	NA	NA	NA	10.115077	7.547538	0.0758724	4.220576	8.733359	14.51579	52.78817	<U+2587><U+2585><U+2581><U+2581><U+2581>
numeric	BA.0.2010	10	0.9746835	NA	NA	NA	NA	NA	11.366633	8.163326	0.1405362	5.171318	9.901115	15.69708	57.18753	<U+2587><U+2585><U+2581><U+2581><U+2581>
numeric	BA.0.2011	9	0.9772152	NA	NA	NA	NA	NA	12.494634	8.797904	0.1697704	5.679900	11.022433	16.68853	58.07456	<U+2587><U+2586><U+2582><U+2581><U+2581>

```
# skim(rgr_raw) - for markdown visualization

kable(skim(msh_raw),"latex", booktabs = T) %>%
  kable_styling(latex_options="scale_down")
```

skim_type	skim_variable	n_missing	complete_rate	character.min	character.max	character.empty	character.n_unique	character.whitespace	numeric.mean	numeric.sd	numeric.p0	numeric.p25	numeric.p50	numeric.p75	numeric.p100	numeric.hist
character	SampleID	0	1.0000000	6	10	0	395	0	NA	NA	NA	NA	NA	NA	NA	NA
character	Species	0	1.0000000	3	4	0	25	0	NA	NA	NA	NA	NA	NA	NA	NA
character	Site	0	1.0000000	3	19	0	23	0	NA	NA	NA	NA	NA	NA	NA	NA
numeric	Height.DBH.Ratio	0	1.0000000	NA	NA	NA	NA	NA	149.7357218	41.4776792	66.2686567	120.0000000	144.6153846	173.8119658	292.0000000	<U+2583><U+2587><U+2585><U+2582><U+2581>
numeric	Estem	0	1.0000000	NA	NA	NA	NA	NA	4056.5447930	3332.2310714	360.8974410	1670.3688930	3158.5020190	5437.7363340	21146.1018900	<U+2587><U+2583><U+2581><U+2581><U+2581>
numeric	Etwig	0	1.0000000	NA	NA	NA	NA	NA	343.4024492	192.6900694	22.9502873	200.9990223	312.1629240	431.5659381	1124.2157920	<U+2586><U+2587><U+2582><U+2581><U+2581>
numeric	Branching.Distance	0	1.0000000	NA	NA	NA	NA	NA	23.9342212	21.2732160	5.1333333	11.0934783	16.1428571	27.3000000	100.0000000	<U+2587><U+2582><U+2581><U+2581><U+2581>
numeric	Twig.Diameter	0	1.0000000	NA	NA	NA	NA	NA	5.5794038	1.8103709	3.0250000	4.4550000	5.1500000	6.2500000	15.5950000	<U+2587><U+2583><U+2581><U+2581><U+2581>
numeric	Twig.Wood.Density	0	1.0000000	NA	NA	NA	NA	NA	0.5018101	0.0773251	0.2697191	0.4558797	0.5115347	0.5538173	0.6759929	<U+2581><U+2583><U+2587><U+2587><U+2582>
numeric	Stem.Wood.Density	0	1.0000000	NA	NA	NA	NA	NA	0.6162300	0.1320942	0.2648281	0.5167250	0.6049548	0.7273644	0.8974172	<U+2581><U+2585><U+2587><U+2587><U+2583>
numeric	Leaf.Mass.Fraction	0	1.0000000	NA	NA	NA	NA	NA	1.2908830	0.6356924	0.2367848	0.8101039	1.1552226	1.6456989	3.5336659	<U+2586><U+2587><U+2583><U+2582><U+2581>
numeric	Leaf.Area	0	1.0000000	NA	NA	NA	NA	NA	59.3640807	78.3130797	3.7444000	18.2803244	29.7848387	61.9248846	474.4577273	<U+2587><U+2581><U+2581><U+2581><U+2581>
numeric	LMA	0	1.0000000	NA	NA	NA	NA	NA	52.5570845	24.2837043	17.4592553	32.9624497	45.3863990	67.7590945	134.2225368	<U+2587><U+2586><U+2583><U+2582><U+2581>
numeric	LCC	0	1.0000000	NA	NA	NA	NA	NA	47.5027803	2.3657203	41.7288000	46.0648000	47.6196000	49.1278000	52.5132000	<U+2582><U+2583><U+2587><U+2586><U+2582>
numeric	LNC	0	1.0000000	NA	NA	NA	NA	NA	2.6920081	0.5066528	1.4219000	2.3377500	2.6778000	3.0049000	4.1779000	<U+2581><U+2586><U+2587><U+2583><U+2581>
numeric	LPC	0	1.0000000	NA	NA	NA	NA	NA	15.3871587	4.8429936	5.5412378	12.0663517	14.4697182	17.8350485	41.3020907	<U+2581><U+2585><U+2587><U+2582><U+2581>
numeric	d15N	0	1.0000000	NA	NA	NA	NA	NA	-2.2298654	2.8404845	-9.0360275	-4.1053924	-2.2042375	-0.7518847	16.1917920	<U+2581><U+2583><U+2587><U+2581><U+2581>
numeric	t.h2	7	0.9822785	NA	NA	NA	NA	NA	0.0250138	0.0090180	0.0007625	0.0191504	0.0235593	0.0302758	0.0580334	<U+2581><U+2587><U+2587><U+2586><U+2582>
numeric	Ks	8	0.9797468	NA	NA	NA	NA	NA	117.0343380	144.7499539	10.9835978	39.2909083	57.4040847	122.2144627	985.8303949	<U+2587><U+2581><U+2581><U+2581><U+2581>
numeric	Ktwig	8	0.9797468	NA	NA	NA	NA	NA	1156.3121313	901.7567951	90.7266105	622.4270895	899.1533593	1331.7824990	7135.0900750	<U+2587><U+2582><U+2581><U+2581><U+2581>
numeric	Huber.Value	4	0.9898734	NA	NA	NA	NA	NA	0.1440426	0.6306790	0.0014640	0.0056178	0.0083976	0.0165698	5.9961641	<U+2587><U+2581><U+2581><U+2581><U+2581>
numeric	X.Lum	8	0.9797468	NA	NA	NA	NA	NA	0.1328261	0.0444313	0.0355744	0.1003895	0.1291688	0.1614926	0.2781733	<U+2582><U+2587><U+2587><U+2587><U+2581>
numeric	VD	8	0.9797468	NA	NA	NA	NA	NA	10521.0525322	9850.2692383	1988.8917790	5195.7487790	6765.9222750	10350.0596500	53249.4432200	<U+2587><U+2581><U+2581><U+2581><U+2581>
numeric	X.Sapwood	4	0.9898734	NA	NA	NA	NA	NA	0.7648435	0.2847063	0.0089190	0.7203337	0.9036817	0.9580712	0.9958401	<U+2581><U+2581><U+2581><U+2582><U+2587>
numeric	d13C	0	1.0000000	NA	NA	NA	NA	NA	-30.2257774	1.4615324	-33.8285092	-31.3569457	-30.4244853	-29.3083104	-25.8225092	<U+2582><U+2587><U+2587><U+2587><U+2581>
numeric	Biomass5	11	0.9721519	NA	NA	NA	NA	NA	6.1811851	7.9143947	0.0319446	1.6859018	3.9305276	7.6296229	86.0563218	<U+2587><U+2581><U+2581><U+2581><U+2581>
numeric	pcent_max.Bio5	11	0.9721519	NA	NA	NA	NA	NA	0.3116330	0.2895216	0.0006399	0.0855931	0.2116297	0.4476936	1.0000000	<U+2587><U+2583><U+2582><U+2581><U+2582>
numeric	Symmetric.Competition	9	0.9772152	NA	NA	NA	NA	NA	2.7314834	3.8865741	0.0221145	0.6096180	1.3246754	3.2648763	27.0583772	<U+2587><U+2581><U+2581><U+2581><U+2581>
numeric	Asymmetric.Competition	9	0.9772152	NA	NA	NA	NA	NA	20.2829284	23.1518057	0.2182907	4.4316511	13.0803780	27.5518557	149.3085902	<U+2587><U+2582><U+2581><U+2581><U+2581>
numeric	Soil.Humidity	0	1.0000000	NA	NA	NA	NA	NA	0.1547939	0.1578224	0.0036667	0.0610000	0.1053333	0.1858333	1.0000000	<U+2587><U+2582><U+2581><U+2581><U+2581>
numeric	pH	130	0.6708861	NA	NA	NA	NA	NA	5.1266792	0.7340680	3.6100000	4.5700000	5.0900000	5.6800000	7.1900000	<U+2583><U+2587><U+2587><U+2587><U+2583>
numeric	Organic.C	130	0.6708861	NA	NA	NA	NA	NA	11.1824289	10.0556779	0.6856316	3.2425755	7.4462036	14.3749874	44.2886336	<U+2587><U+2583><U+2581><U+2582><U+2581>
numeric	N	130	0.6708861	NA	NA	NA	NA	NA	0.4550739	0.4793418	0.0076733	0.1025588	0.2692148	0.5848671	2.1087924	<U+2587><U+2582><U+2581><U+2581><U+2581>
numeric	P	130	0.6708861	NA	NA	NA	NA	NA	32.2434455	39.6926773	2.1562818	12.0228980	18.1015122	33.9273519	322.5392749	<U+2587><U+2581><U+2581><U+2581><U+2581>
numeric	K	130	0.6708861	NA	NA	NA	NA	NA	138.2512003	95.1751850	24.2527806	67.2212574	112.7999536	189.7025184	533.1378680	<U+2587><U+2585><U+2581><U+2581><U+2581>
numeric	Ca	130	0.6708861	NA	NA	NA	NA	NA	207.4988264	2312.2158636	116.9189371	505.7524772	1351.8078700	2564.6472360	14134.2452000	<U+2587><U+2581><U+2581><U+2581><U+2581>
numeric	Mg	130	0.6708861	NA	NA	NA	NA	NA	107.0133431	93.3185973	9.4122999	40.2245695	78.6918416	149.2825380	615.6952603	<U+2587><U+2583><U+2581><U+2581><U+2581>
numeric	Soil.Depth	82	0.7924051	NA	NA	NA	NA	NA	33.1948882	14.8095941	3.0000000	21.0000000	30.0000000	46.0000000	55.0000000	<U+2582><U+2586><U+2587><U+2583><U+2587>
numeric	Slope	0	1.0000000	NA	NA	NA	NA	NA	8.6543367	10.1938941	0.0000000	0.0000000	5.3000000	15.0000000	39.6000000	<U+2587><U+2583><U+2582><U+2581><U+2581>
numeric	North.Aspect	0	1.0000000	NA	NA	NA	NA	NA	-0.1690550	0.4854480	-1.0000000	-0.5149596	0.0000000	0.0000000	0.9902681	<U+2583><U+2581><U+2581><U+2587><U+2581>
numeric	East.Aspect	0	1.0000000	NA	NA	NA	NA	NA	0.0594598	0.5074686	-0.9993908	0.0000000	0.0000000	0.2419219	1.0000000	<U+2581><U+2582><U+2581><U+2587><U+2581><U+2582>
numeric	Elevation	0	1.0000000	NA	NA	NA	NA	NA	142.4911392	76.4742868	40.0000000	81.5000000	146.0000000	200.0000000	350.0000000	<U+2586><U+2587><U+2587><U+2585><U+2581>
numeric	Summer.Max	0	1.0000000	NA	NA	NA	NA	NA	27.1680759	1.6051104	24.1700000	26.1700000	26.6700000	27.7500000	32.1700000	<U+2582><U+2587><U+2587><U+2582><U+2581>
numeric	Fall.Min	0	1.0000000	NA	NA	NA	NA	NA	-10.2153418	1.2660912	-14.6700000	-10.8300000	-10.5000000	-9.5000000	-8.0000000	<U+2581><U+2581><U+2587><U+2587><U+2586>
numeric	Mean.Temp	0	1.0000000	NA	NA	NA	NA	NA	13.7220000	0.5813513	12.3100000	13.3200000	13.8300000	14.1000000	14.6700000	<U+2581><U+2583><U+2583><U+2585><U+2587><U+2585>
numeric	X.Canopy.Opening	0	1.0000000	NA	NA	NA	NA	NA	9.1949065	7.9404999	1.0600000	4.7650000	6.4000000	10.5050000	57.0100000	<U+2587><U+2581><U+2581><U+2581><U+2581>
numeric	Tree.Age	6	0.9848101	NA	NA	NA	NA	NA	20.8226221	10.3881823	5.0000000	14.0000000	17.0000000	26.0000000	66.0000000	<U+2587><U+2585><U+2582><U+2581><U+2581>
numeric	Tree.Height	0	1.0000000	NA	NA	NA	NA	NA	423.4204990	131.4480550	170.0000000	328.5000000	404.5994598	509.0000000	932.0900604	<U+2581><U+2585><U+2587><U+2585><U+2581>
numeric	julian.date.2011	0	1.0000000	NA	NA	NA	NA	NA	194.3518987	12.1870038	171.0000000	186.0000000	193.0000000	204.0000000	218.0000000	<U+2583><U+2587><U+2583><U+2582><U+2583>
numeric	Branch.Height	0	1.0000000	NA	NA	NA	NA	NA	3.0436709	1.0307631	1.2000000	2.3000000	2.9000000	3.6500000	6.8000000	<U+2581><U+2585><U+2587><U+2585><U+2582><U+2581>
numeric	SRL	242	0.3873418	NA	NA	NA	NA	NA	105.8022774	60.1145386	14.1784286	64.2643965	92.0143246	143.7505535	288.1376200	<U+2587><U+2587><U+2587><U+2585><U+2582><U+2581>
numeric	Root.Wood.Density	196	0.5037975	NA	NA	NA	NA	NA	0.5683471	0.0990170	0.2954545	0.5098836	0.5695090	0.6401440	0.8468866	<U+2581><U+2581><U+2585><U+2587><U+2586><U+2581>
numeric	Twig.branching.angle	100	0.7468354	NA	NA	NA	NA	NA	53.7748704	16.9901910	0.0000000	46.8824125	54.9505000	64.5750500	102.9261500	<U+2581><U+2581><U+2582><U+2587><U+2585><U+2581>
numeric	Hmax	0	1.0000000	NA	NA	NA	NA	NA	20.2582278	9.2161936	5.0000000	12.0000000	25.0000000	25.0000000	35.0000000	<U+2586><U+2585><U+2582><U+2587><U+2585>
numeric	Shade.Tolerance	37	0.9063291	NA	NA	NA	NA	NA	2.7761732	1.1405385	1.2100000	1.5600000	2.5900000	3.5600000	4.7600000	<U+2587><U+2586><U+2585><U+2585><U+2583>
numeric	Drought.Tolerance	37	0.9063291	NA	NA	NA	NA	NA	2.4795531	0.7042697	1.5000000	1.7700000	2.3800000	2.9200000	4.0000000	<U+2587><U+2583><U+2586><U+2582><U+2582>
numeric	WaterLogging.Tolerance	37	0.9063291	NA	NA	NA	NA	NA	1.7887989	0.7711961	1.0000000	1.0700000	1.5000000	2.5000000	3.3700000	<U+2587><U+2582><U+2582><U+2582><U+2583><U+2582>
numeric	DBH.DGH	0	1.0000000	NA	NA	NA	NA	NA	0.6706239	0.1062622	0.2857143	0.6220754	0.6911765	0.7430852	0.9714286	<U+2581><U+2582><U+2587><U+2587><U+2581>
numeric	Leaf.Area.Ratio	0	1.0000000	NA	NA	NA	NA	NA	266.3885148	134.2717209	32.1518987	165.3256050	243.9272727	349.2873753	756.6600000	<U+2586><U+2585><U+2587><U+2583><U+2581>
numeric	Branch.Dry.Mass	0	1.0000000	NA	NA	NA	NA	NA	5.5823852	4.0655544	1.0610000	2.9790000	4.4588618	6.6222105	30.6862626	<U+2587><U+2582><U+2581><U+2581><U+2581>
numeric	Leaf.Thickness	0	1.0000000	NA	NA	NA	NA	NA	0.1349420	0.0369056	0.0628000	0.1082000	0.1278000	0.1510000	0.2442000	<U+2582><U+2587><U+2587><U+2585><U+2582><U+2581>
numeric	Leaf.Density	0	1.0000000	NA	NA	NA	NA	NA	0.3791845	0.0998209	0.1435794	0.3088997	0.3704841	0.4436658	0.6931874	<U+2581><U+2582><U+2587><U+2587><U+2583><U+2581>
numeric	C.N	0	1.0000000	NA	NA	NA	NA	NA	18.2776260	3.7301501	10.2515895	15.7189855	17.7223137	20.3719408	35.2219038	<U+2583><U+2583><U+2587><U+2583><U+2581><U+2581>
numeric	C.P	0	1.0000000	NA	NA	NA	NA	NA	3.3536266	0.9963332	1.1588421	2.6520406	3.2477076	3.9828974	8.3173836	<U+2583><U+2583><U+2587><U+2583><U+2581><U+2581>
numeric	N.P	0	1.0000000	NA	NA	NA	NA	NA	0.1864329	0.0514411	0.0550209	0.1513173	0.1839044	0.2180884	0.4173616	<U+2582><U+2582><U+2587><U+2587><U+2586><U+2581><U+2581>
numeric	Ki	8	0.9797468	NA	NA	NA	NA	NA	811.5405105	817.6039150	154.6714211	349.4009763	466.0318465	730.0596264	4545.3497270	<U+2587><U+2581><U+2581><U+2581><U+2581>
numeric	Kleaf	8	0.9797468	NA	NA	NA	NA	NA	1.4535362	2.0635095	0.0055578	0.4866632	0.9339347	1.6383879	22.4651990	<U+2587><U+2581><U+2581><U+2581><U+2581>
numeric	Biomass1	11	0.9721519	NA	NA	NA	NA	NA	1.2258952	1.9751131	0.0008279	0.2439121	0.5513097	1.2903255	18.1522231	<U+2587><U+2581><U+2581><U+2581><U+2581>
numeric	pcent_max.Bio1	11	0.9721519	NA	NA	NA	NA	NA	0.2469674	0.2742624	0.0026217	0.0527320	0.1321388	0.3210158	1.0000000	<U+2587><U+2582><U+2581><U+2581><U+2581>
numeric	Total.Transmitted.Light	0	1.0000000	NA	NA	NA	NA	NA	5.5416696	5.1148204	0.6300000	2.6450000	3.8200000	6.1400000	36.1800000	<U+2587><U+2581><U+2581><U+2581><U+2581>

```
# skim(msh_raw) - for markdown visualization
```

Including Plots

Might be best to just drop some the columns trait values with a large % of missing data.

For the trait data, I will simply remove those with a majority of missing cases before running the imputation - I am thinking of using a cut of less than 50%.

```
# removing cases with a bunch of zeroes
MSH.50 <- msh_raw[,names(which(sapply(colnames(msh_raw), function (x) {sum(is.na(msh_raw[,x]))/nrow(msh
MSH.5 <- msh_raw[,names(which(sapply(colnames(msh_raw), function (x) {sum(is.na(msh_raw[,x]))/nrow(msh_
# I only want columns that are numeric
numeric_columns_msh <- which(apply(MSH.50,2, function (x) !is.na(mean(as.numeric(x), na.rm = TRUE))))
numeric_columns_rgr <- which(apply(rgr_raw,2, function (x) !is.na(mean(as.numeric(x), na.rm = TRUE))))
# I read somewhere that p>5 does not change the overall values significantly So, I will stick to that n
MSH.IP.MICE <- mice(data.matrix(MSH.50[,numeric_columns_msh]),
                    printFlag = FALSE, method = "pmm", m = 5, seed = 300)
```

Let's combine the imputed data - just to check if a regression would give different results.

```
library(abind)
MSH.IP.List <- list()
m_length <- MSH.IP.MICE$m
for (i in 1:m_length){
    MSH.IP.List[[i]] <- mice::complete(MSH.IP.MICE, action = i)}
MSH.IP.Final <- data.frame(apply(abind(MSH.IP.List, along=3), c(1,2), mean))
```

Now, I will run the imputation on these data. Since I have removed the columns with a large amount of missing cases, the imputation should work. If not, I will set a more stringent condition.

```
# fit a lm and see if results are comparable between mice output and raw data
summary(with(data = MSH.IP.Final, exp = lm(Tree.Height ~ Soil.Depth + Biomass1 + Huber.Value  + pH)))
```

```
##
## Call:
## lm(formula = Tree.Height ~ Soil.Depth + Biomass1 + Huber.Value +
##     pH)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -261.03  -97.07  -16.84   88.79  522.04
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 366.6112    54.4005   6.739 5.74e-11 ***
## Soil.Depth    0.8010     0.4709   1.701 0.089781 .
## Biomass1     11.7975     3.4656   3.404 0.000732 ***
## Huber.Value -15.9981    10.9075  -1.467 0.143261
## pH            3.4535     9.8303   0.351 0.725540
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 129.2 on 390 degrees of freedom
## Multiple R-squared:  0.04325,    Adjusted R-squared:  0.03343
## F-statistic: 4.407 on 4 and 390 DF,  p-value: 0.001701
```

```
# fit a lm and see if results are comparable between mice output and raw data
summary(lm(Tree.Height ~ Soil.Depth + Biomass1 + Huber.Value + pH, data = msh_raw))
```

```
##
## Call:
## lm(formula = Tree.Height ~ Soil.Depth + Biomass1 + Huber.Value +
##     pH, data = msh_raw)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -275.02 -100.11  -25.28   87.24  527.31
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 376.5703    60.2157   6.254 1.67e-09 ***
## Soil.Depth    0.4151     0.5481   0.757   0.4495
## Biomass1     20.7473     4.6604   4.452 1.27e-05 ***
## Huber.Value 313.5464   135.3190   2.317   0.0213 *
## pH            1.9247    11.0954   0.173   0.8624
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 130.9 on 256 degrees of freedom
##   (134 observations deleted due to missingness)
```

```
## Multiple R-squared:  0.09589,    Adjusted R-squared:  0.08176
## F-statistic: 6.788 on 4 and 256 DF,  p-value: 3.307e-05
```

The regression coefficients do vary slightly. But the direction and significance of these relationships are near the same.

```
# how do the combined and imputed datasets compare?
xyplot(MSH.IP.MICE, Soil.Depth ~ Soil.Humidity|.imp, pch = 20, cex = 1.4)
```


\includegraphics[ ]{data-exploration-cleaning_files/figure-latex/unnamed-chunk-6-1}

```
xyplot(MSH.IP.MICE, Soil.Depth ~ pH|.imp, pch = 20, cex = 1.4)
```


\includegraphics[ ]{data-exploration-cleaning_files/figure-latex/unnamed-chunk-6-2}

```
#The imputed data seems to match that of the raw data. Now, how do the combined imputed data compare to
merge_imputations(msh_raw, MSH.IP.MICE, summary =  "dens")
```


\includegraphics[ ]{data-exploration-cleaning_files/figure-latex/unnamed-chunk-6-3}

There is strong overlap between the mean and merged values. As such, I can be confident that the merged data set is representative of individual imputations.

Estimate Growth Rate

There are three metrics we can use to measure growth rates. They are: * Basal Area Relative Growth Rate * Biomass Relative Growth Rate

```
# makes more sense to name the rows by their sample ID as a unique identifier
MSH.IP.Final$SampleID <- rgr_raw[,1]

MSH.RGR_IMP <- merge(rgr_raw,MSH.IP.Final, by = "SampleID")
MSH.RGR <- merge(rgr_raw,MSH.50, by = "SampleID")
# now for growth rate parameters
BAI_GR <- with(data = MSH.RGR_IMP, exp = ((BA.0.2011 - BA.0.2006)/5))
BIO_Gain <- with(data = MSH.RGR_IMP,
    exp = ((BA.0.2011*Stem.Wood.Density*Tree.Height*3)-(BA.0.2006*Stem.Wood.Density*(Tree.Height/Tree.Ag
BIO_2006 <- with(data = MSH.RGR_IMP,
                 exp = (BA.0.2006*Stem.Wood.Density*(Tree.Height/Tree.Age*(Tree.Age-6))*3))
BIO_GR <- BIO_Gain/((BIO_2006+1)*5)
# wanna place species and site at the start of the data frame
site_species <- c("SampleID", "Site", "Species", "Porosity")
site_species_no <- which(colnames(MSH.50)%in%site_species)
site_BA <- c(colnames(rgr_raw)[-1], "SampleID")
site_BA_no <- which(colnames(MSH.RGR_IMP)%in%site_BA)

RGR_MSH_Final <- data.frame(MSH.50[,site_species_no], BAI_GR = BAI_GR,
                            BIO_GR = BIO_GR,
                            MSH.RGR_IMP[,-site_BA_no])

# now for growth rate parameters
```

```
BAI_GR_NA <- with(data = MSH.RGR, exp = ((BA.0.2011 - BA.0.2006)/5))
BIO_Gain_NA <- with(data = MSH.RGR, exp = ((BA.0.2011*Stem.Wood.Density*Tree.Height*3) - (BA.0.2006*Ste
BIO_2006_NA <- with(data = MSH.RGR, exp = (BA.0.2006*Stem.Wood.Density*(Tree.Height/Tree.Age*(Tree.Age-
BIO_GR_NA <- BIO_Gain_NA/((BIO_2006_NA+1)*5)

RGR_MSH_NA_Final <- data.frame(MSH.50[,site_species_no], BAI_GR = BAI_GR_NA,
                           BIO_GR = BIO_Gain_NA,
                           MSH.50[,numeric_columns_msh])


write_csv(RGR_MSH_Final,here("data/RGR_MSH.csv"))
write_csv(RGR_MSH_NA_Final,here("data/RGR_MSH_NA.csv"))
```