Tree Data Analysis (ID issue fixed)

Tahmidul Islam

12/16/2020

```
library(dplyr)
library(ggplot2)
library(kableExtra)
library(tidyr)
```

Tree Data

This dataset contains features of trees (height, width, etc.) from 3 sites: Brighton, Chase Stream and Lily Bay. There are 8 species in this sample.

Table: Data overview.

```
tree <- read.csv("E:/R Project/BayesFDA/data/tree/tree.csv")
tree <- tree %>% dplyr::select(Site, ID.Code, Tree, Rep, Sp, Year, Height) %>%
  filter(!is.na(Height)) %>%
  mutate(id = paste0(Site, ID.Code, Rep, Tree), logHeight = log(Height))
head(tree) %>% kable()
```

Site	ID.Code	Tree	Rep	Sp	Year	Height	id	logHeight
Brighton	S-BS-Mlt	1	1	BS	5	4.5	BrightonS-BS-Mlt11	1.504077
Brighton	S-BS-Mlt	1	1	BS	10	10.4	BrightonS-BS-Mlt11	2.341806
Brighton	S-BS-Mlt	2	1	BS	5	6.8	BrightonS-BS-Mlt12	1.916923
Brighton	S-BS-Mlt	2	1	BS	10	13.1	BrightonS-BS-Mlt12	2.572612
Brighton	S-BS-Mlt	4	1	BS	5	4.4	BrightonS-BS-Mlt14	1.481604
Brighton	S-BS-Mlt	4	1	BS	10	10.3	BrightonS-BS-Mlt14	2.332144

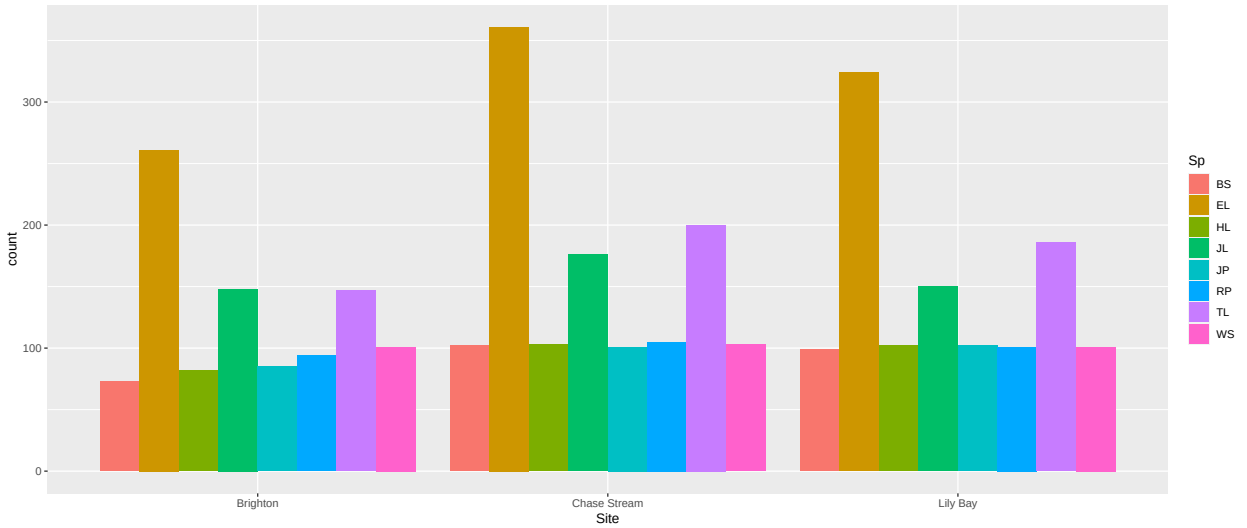
Table: Distribution of site and species.

```
sumDf <- tree %>% group_by(Site, Sp, id) %>% summarize (n = n()) %>% dplyr::select(Site, Sp)

sumDf %>% tally() %>% spread(Sp, n) %>% kable()
```

Site	BS	EL	HL	JL	JP	RP	TL	WS
Brighton	73	261	82	148	85	94	147	101
Chase Stream	102	361	103	176	101	105	200	103
Lily Bay	99	324	102	150	102	101	186	101

```
ggplot(sumDf, aes(x = Site, fill = Sp)) + geom_bar(position="dodge", stat="count")
```

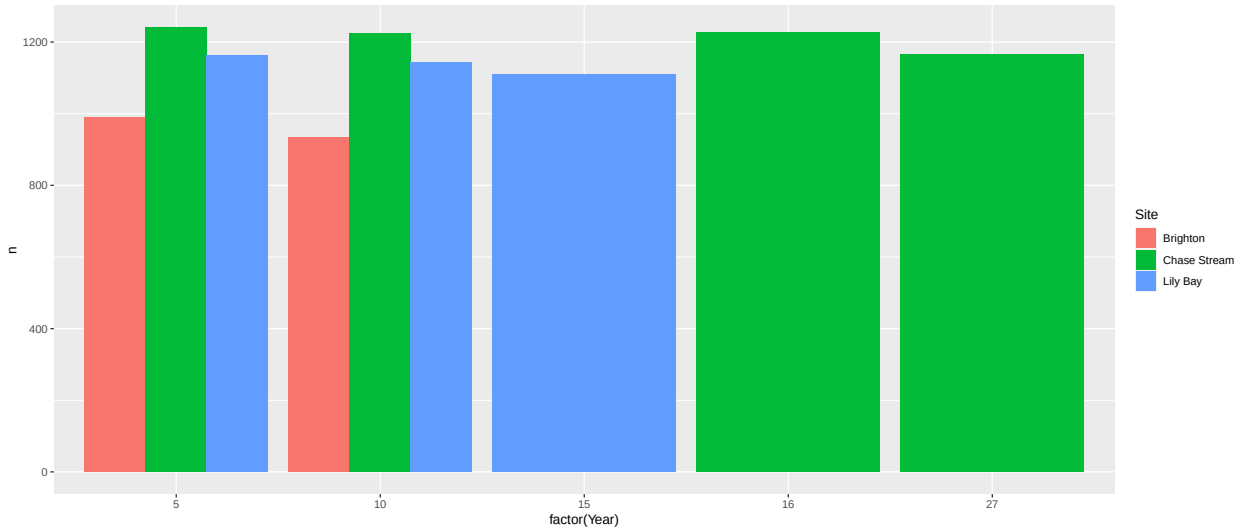


The measurements are collected few years apart: at 5, 10, 15, 16 and 27 years.

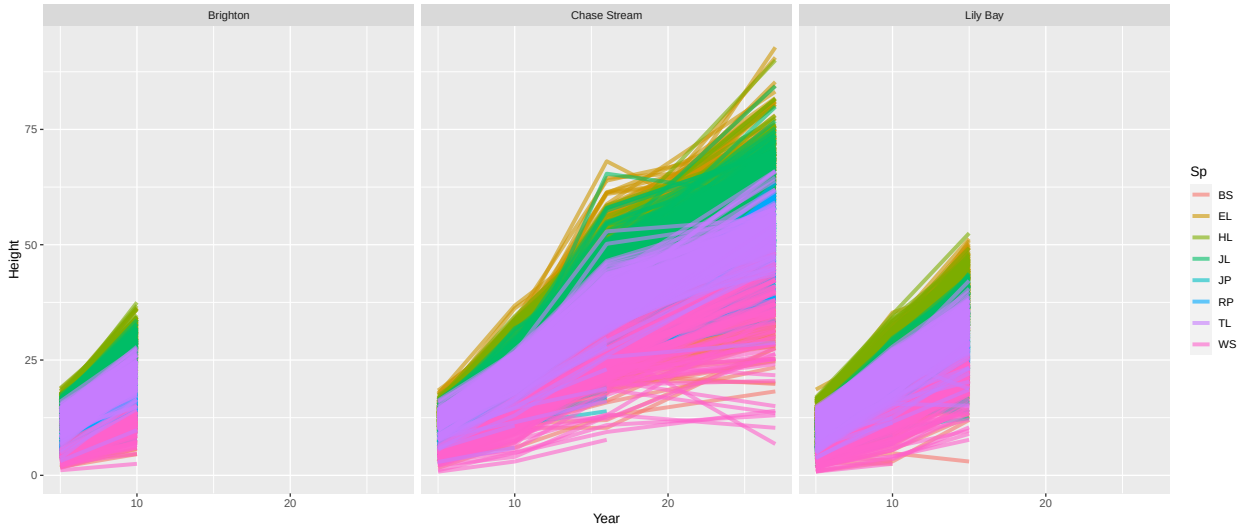
```
sumDf <- tree %>% group_by(Site, Year)  %>%  summarize(n = n())
sumDf %>% kable()
```

Site	Year	n
Brighton	5	989
Brighton	10	933
Chase Stream	5	1240
Chase Stream	10	1224
Chase Stream	16	1226
Chase Stream	27	1166
Lily Bay	5	1162
Lily Bay	10	1144
Lily Bay	15	1109

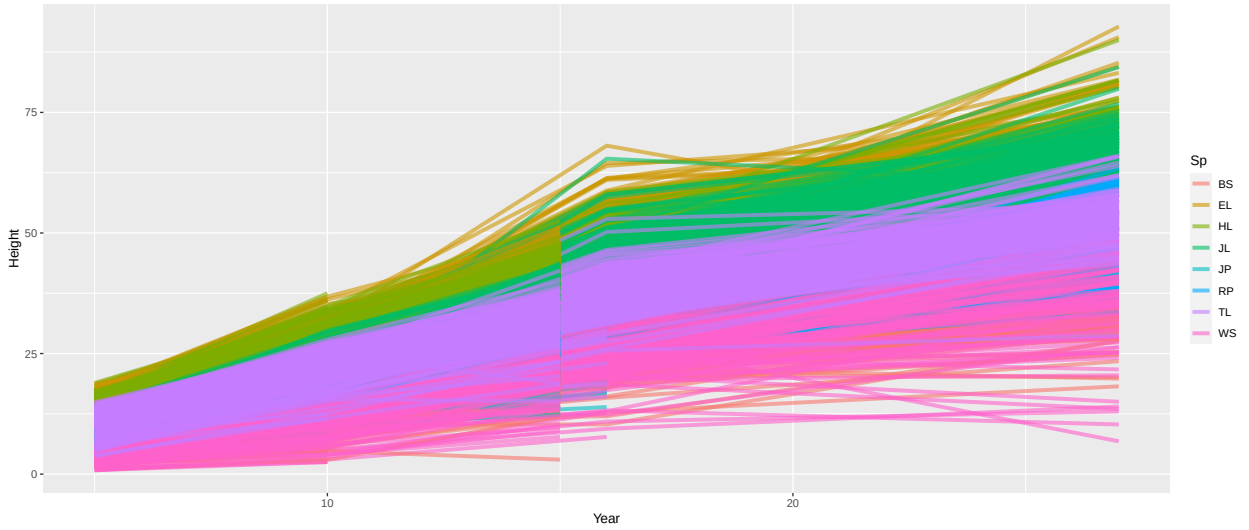
```
ggplot(sumDf, aes(x = factor(Year), y = n, fill = Site)) + geom_col(position = 'dodge', stat = 'count')
```




```
ggplot(data = tree, aes(x = Year, y = Height, col = Sp, group = id)) +
  geom_line(size = 1.5, alpha = .6) + facet_wrap(~Site)
```




```
ggplot(data = tree, aes(x = Year, y = Height, col = Sp, group = id)) +
  geom_line(size = 1.5, alpha = .6)
```

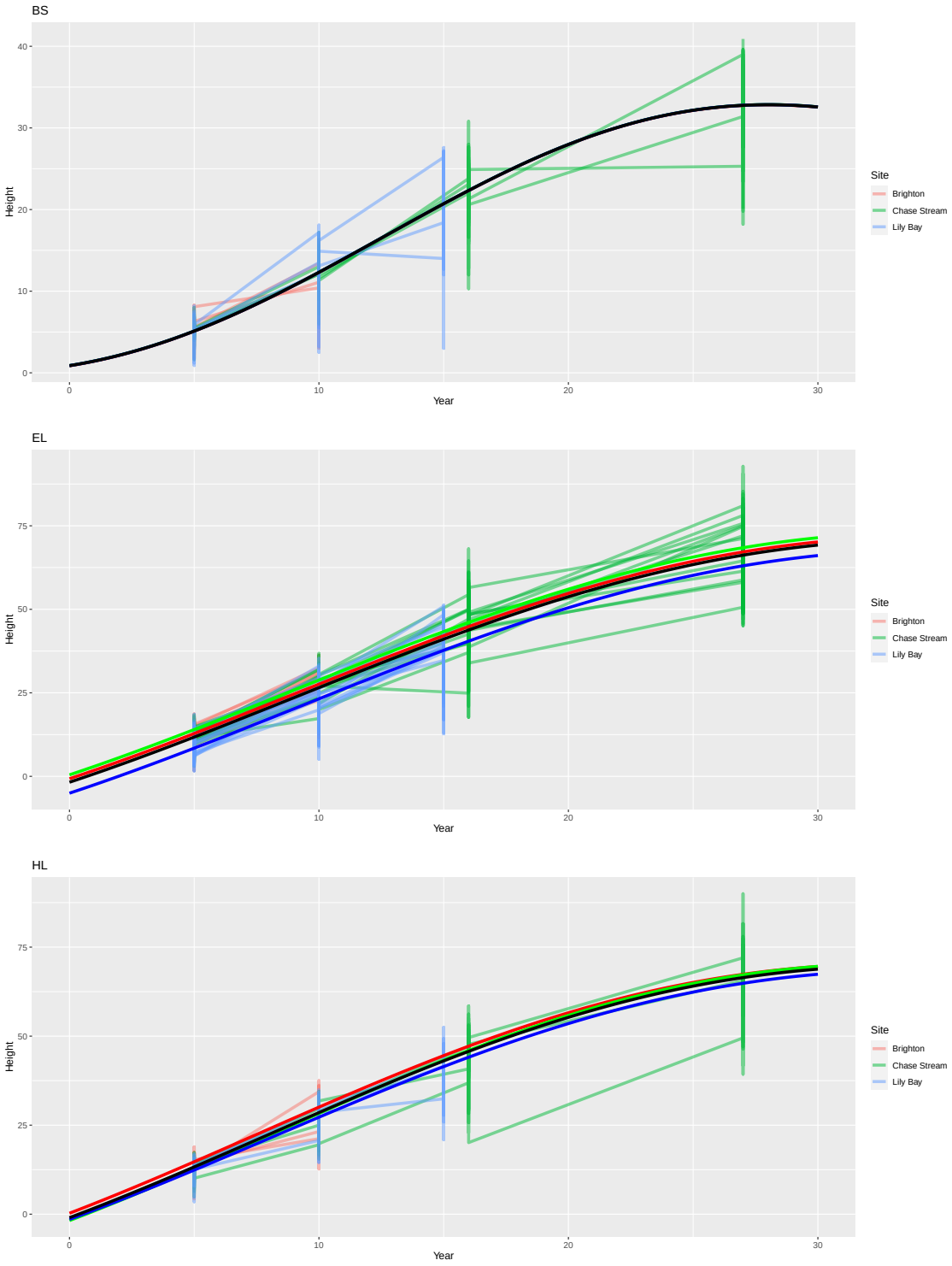


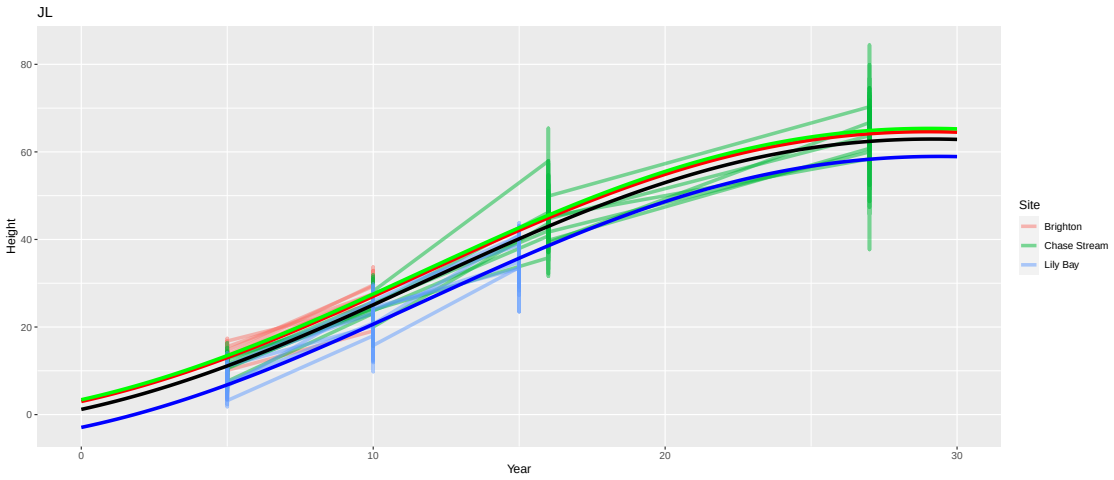
GP model: Site Specific Mean Height Estimation

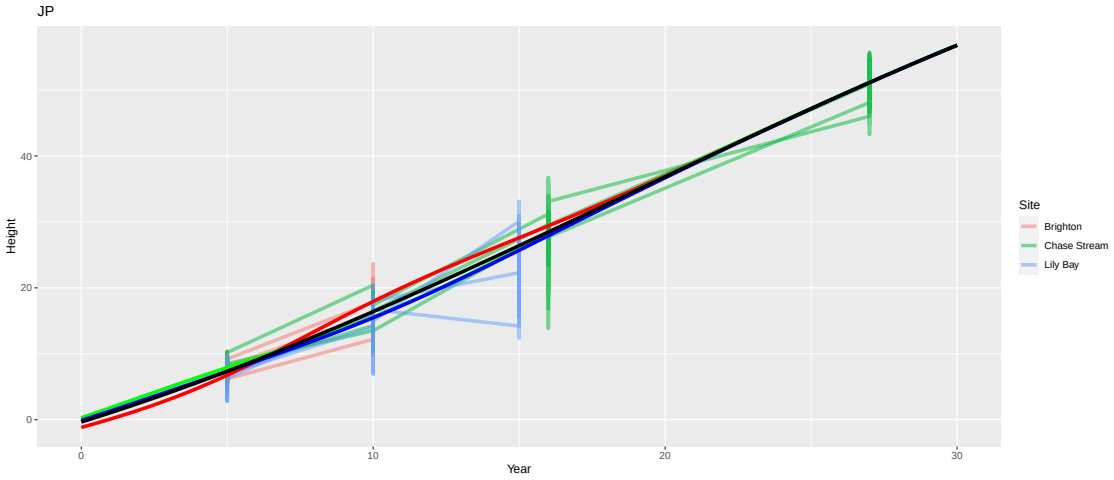
Parameter estimates

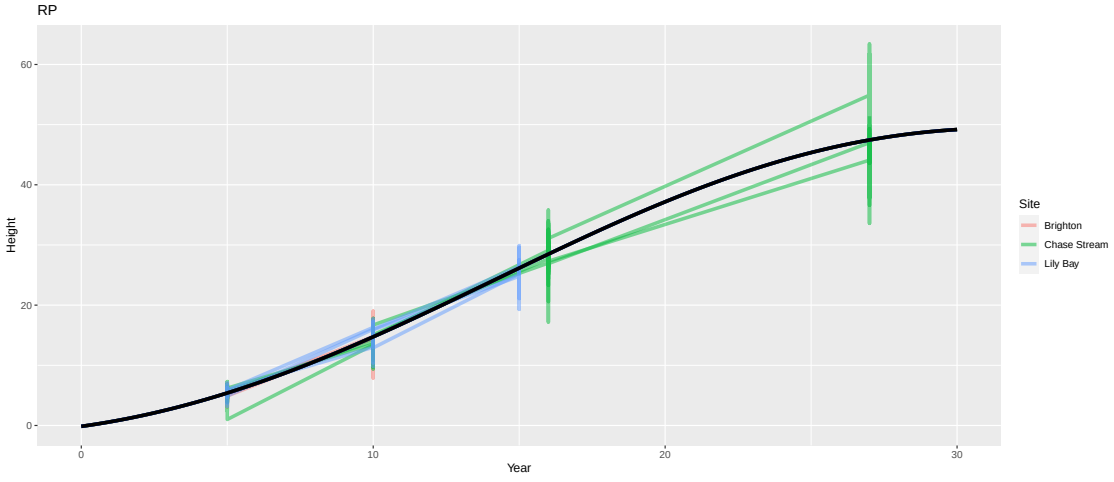
	Mean		Subject		Site		
	l^2	σ^2	l^2	σ^2	l^2	σ^2	σ^2
BS	0.85	3.46	1.20	0.00	0.07	0.00	0.09
EL	2.29	9.78	0.46	0.01	0.12	0.01	0.06
HL	1.90	5.84	0.55	0.00	1.38	0.01	0.08
JL	0.98	2.66	0.33	0.01	0.36	0.02	0.04
JP	3.92	11.38	0.07	0.00	18.92	0.00	0.04
RP	1.06	4.13	0.13	0.03	4.46	0.00	0.02
TL	1.12	3.09	1.68	0.00	0.01	0.00	0.06
WS	0.69	2.11	0.30	0.02	0.03	0.00	0.13

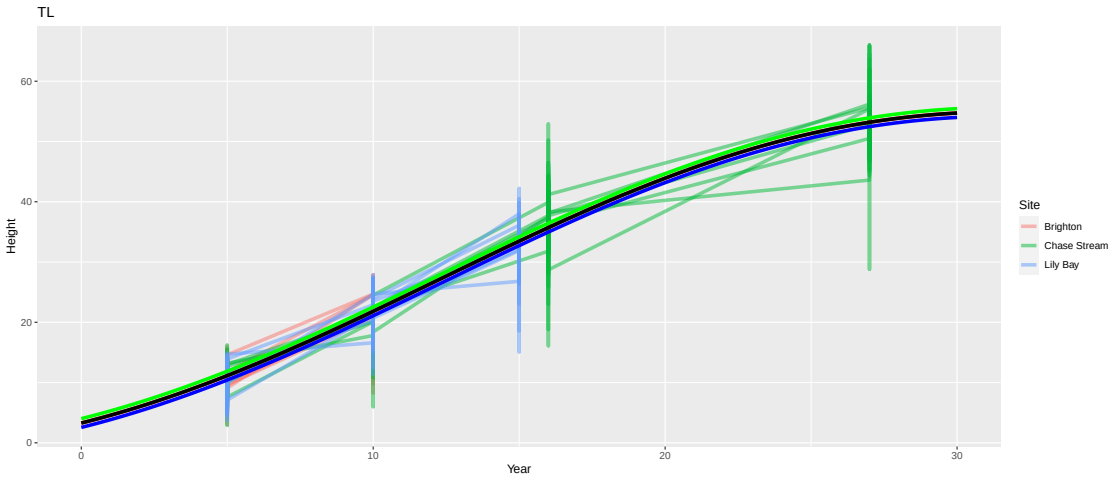
Plots of site specific estimation

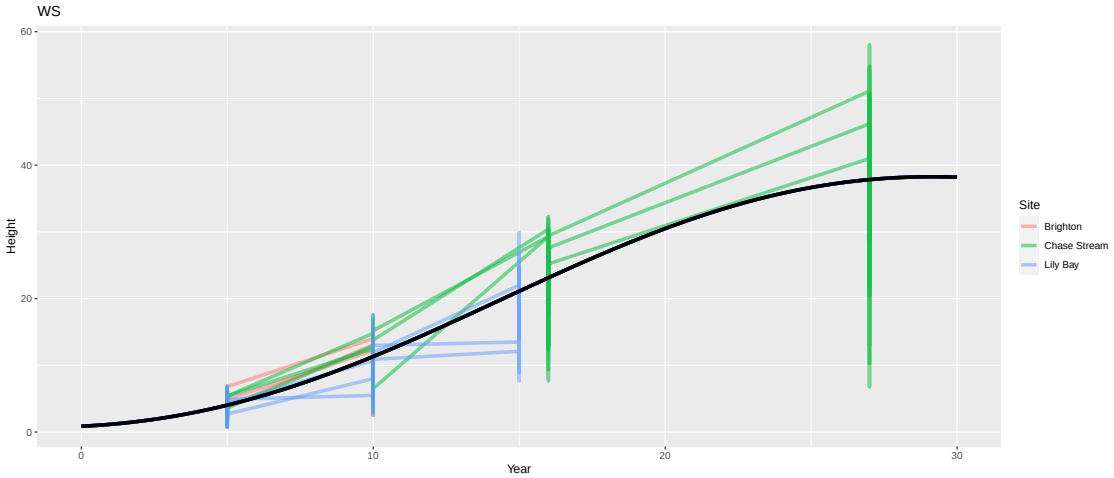



JL
Height
Year
Site
Brighton
Chase Stream
Lily Bay


JP
Height
Year
Site
Brighton
Chase Stream
Lily Bay


RP
Height
Year
Site
Brighton
Chase Stream
Lily Bay


TL
Height
Year
0
20
40
60
0
10
20
30
Site
Brighton
Chase Stream
Lily Bay


WS
Height
Year
0
20
40
60
0
10
20
30
Site
Brighton
Chase Stream
Lily Bay