

uhuru dataset

Luna L Sanchez Reyes

2022-10-04

1. Describing the data that we are using

We are using the dataset from this study

Add a picture of an Acacia

2. Read the data set into R

2.1 Get the working directory of my R chunks The working directory of an Rmd file is not the same as the working directory of the Rstudio project you are in. We can check that by running the following.

How do I test where the computer is?

```
r_proj_wd <- "/Users/lunasare/Desktop/data-science-course/fall-2022"
r_chunk_wd <- getwd()
r_proj_wd == r_chunk_wd
```

```
## [1] FALSE
```

```
r_chunk_wd
```

```
## [1] "/Users/lunasare/Desktop/data-science-course/fall-2022/documents"
```

As you can see, the working directory from my R project and the one used in my R chunks in an Rmd file are not the same.

The working directory of an R chunk is equal to the path where the Rmd file is in, in my case, it is `r_chunk_wd`. This is nested withing my project directory, keep this in mind for when you have to provide the path of a file ;)

```
acacia <- read.csv(file = "../data-raw/ACACIA_DREPANOLOBIUM_SURVEY.txt", sep = "\t")
acacia
```

2.2 Read the data set into R and create an object

##	SURVEY	YEAR	SITE	BLOCK	TREATMENT	PLOT	ID	HEIGHT	AXIS1	AXIS2	CIRC
## 1	1	2012	SOUTH	1	TOTAL	S1TOTAL	581	2.25	2.75	2.15	20.0
## 2	1	2012	SOUTH	1	TOTAL	S1TOTAL	582	2.65	4.10	3.90	28.0
## 3	1	2012	SOUTH	1	TOTAL	S1TOTAL	3111	1.5	1.70	0.85	17.0
## 4	1	2012	SOUTH	1	TOTAL	S1TOTAL	3112	2.01	1.80	1.60	12.0
## 5	1	2012	SOUTH	1	TOTAL	S1TOTAL	3113	1.75	1.84	1.42	13.0
## 6	1	2012	SOUTH	1	TOTAL	S1TOTAL	3114	1.65	1.62	0.85	15.0
## 7	1	2012	SOUTH	1	TOTAL	S1TOTAL	3115	1.2	1.95	0.90	9.0
## 8	1	2012	SOUTH	1	TOTAL	S1TOTAL	3199	1.45	2.00	1.75	12.2
## 9	1	2012	SOUTH	1	MESO	S1MESO	941	1.87	2.15	1.82	13.0
## 10	1	2012	SOUTH	1	MESO	S1MESO	942	2.38	5.55	4.82	35.0
## 11	1	2012	SOUTH	1	MESO	S1MESO	943	2.58	4.90	4.24	24.0

## 12	1	2012	SOUTH	1	MESO	S1MESO	944	2.65	3.75	3.10	27.0
## 13	1	2012	SOUTH	1	MESO	S1MESO	946	2.35	2.34	2.05	20.0
## 14	1	2012	SOUTH	1	MESO	S1MESO	947	1.88	2.10	1.85	28.0
## 15	1	2012	SOUTH	1	MESO	S1MESO	3116	2.32	3.05	2.63	30.0
## 16	1	2012	SOUTH	1	MESO	S1MESO	3117	2.39	2.21	2.10	13.0
## 17	1	2012	SOUTH	1	MESO	S1MESO	3118	2.2	1.80	1.50	10.0
## 18	1	2012	SOUTH	1	MESO	S1MESO	3119	1.05	0.90	0.55	8.0
## 19	1	2012	SOUTH	1	MESO	S1MESO	3120	2	1.25	1.20	10.0
## 20	1	2012	SOUTH	1	MESO	S1MESO	3131	1.28	1.14	1.00	10.0
## 21	1	2012	SOUTH	2	OPEN	S2OPEN	341	dead	NA	NA	NA
## 22	1	2012	SOUTH	2	TOTAL	S2TOTAL	3178	1.4	2.50	2.15	18.0
## 23	1	2012	SOUTH	2	TOTAL	S2TOTAL	101	1.9	3.31	2.65	15.0
## 24	1	2012	SOUTH	2	TOTAL	S2TOTAL	102	1.75	2.70	2.55	16.0
## 25	1	2012	SOUTH	2	TOTAL	S2TOTAL	103	1.8	2.75	2.30	16.0
## 26	1	2012	SOUTH	2	TOTAL	S2TOTAL	104	2.7	4.05	4.00	35.2
## 27	1	2012	SOUTH	2	TOTAL	S2TOTAL	105	2.02	2.85	1.49	17.0
## 28	1	2012	SOUTH	2	TOTAL	S2TOTAL	108	1.9	3.10	2.85	19.0
## 29	1	2012	SOUTH	2	TOTAL	S2TOTAL	109	1.85	2.45	1.90	19.0
## 30	1	2012	SOUTH	2	TOTAL	S2TOTAL	110	1.65	1.90	1.54	17.0
## 31	1	2012	SOUTH	2	TOTAL	S2TOTAL	111	1.4	2.35	1.45	14.0
## 32	1	2012	SOUTH	2	TOTAL	S2TOTAL	113	2.5	3.25	2.30	22.0
## 33	1	2012	SOUTH	2	TOTAL	S2TOTAL	115	2.05	5.40	4.50	33.0
## 34	1	2012	SOUTH	2	TOTAL	S2TOTAL	116	2.26	3.50	3.10	33.0
## 35	1	2012	SOUTH	2	TOTAL	S2TOTAL	117	2.13	2.40	2.30	20.0
## 36	1	2012	SOUTH	2	TOTAL	S2TOTAL	118	1.8	3.15	2.55	22.0
## 37	1	2012	SOUTH	2	TOTAL	S2TOTAL	1211	1.85	2.00	2.27	20.0
## 38	1	2012	SOUTH	2	TOTAL	S2TOTAL	1212	1.5	2.15	1.80	15.0
## 39	1	2012	SOUTH	2	TOTAL	S2TOTAL	1213	1.87	2.34	2.05	13.0
## 40	1	2012	SOUTH	2	TOTAL	S2TOTAL	1214	1.58	1.28	0.75	11.0
## 41	1	2012	SOUTH	2	TOTAL	S2TOTAL	1215	2.05	2.10	1.75	17.0
## 42	1	2012	SOUTH	2	TOTAL	S2TOTAL	1216	1.75	2.45	3.28	16.0
## 43	1	2012	SOUTH	2	TOTAL	S2TOTAL	1217	1.49	1.50	1.45	13.0
## 44	1	2012	SOUTH	2	TOTAL	S2TOTAL	1218	1.28	2.00	0.90	10.0
## 45	1	2012	SOUTH	2	TOTAL	S2TOTAL	1219	1.49	2.35	1.65	13.0
## 46	1	2012	SOUTH	2	TOTAL	S2TOTAL	1220	1.07	1.20	0.95	11.0
## 47	1	2012	SOUTH	2	TOTAL	S2TOTAL	1231	1.48	1.25	1.20	9.0
## 48	1	2012	SOUTH	2	TOTAL	S2TOTAL	1232	1.25	1.25	0.90	10.0
## 49	1	2012	SOUTH	2	TOTAL	S2TOTAL	1233	1.41	1.41	1.40	14.0
## 50	1	2012	SOUTH	2	TOTAL	S2TOTAL	1234	1.6	1.60	1.30	13.0
## 51	1	2012	SOUTH	2	TOTAL	S2TOTAL	1235	1.2	1.20	1.30	14.0
## 52	1	2012	SOUTH	2	TOTAL	S2TOTAL	1236	1.49	1.49	1.20	8.0
## 53	1	2012	SOUTH	2	TOTAL	S2TOTAL	1237	1.5	1.50	1.50	14.0
## 54	1	2012	SOUTH	2	TOTAL	S2TOTAL	1238	1.65	1.65	2.00	20.0
## 55	1	2012	SOUTH	2	TOTAL	S2TOTAL	1239	1.13	1.13	1.20	10.0
## 56	1	2012	SOUTH	2	TOTAL	S2TOTAL	1240	1.25	1.25	0.90	10.0
## 57	1	2012	SOUTH	2	TOTAL	S2TOTAL	1251	1.1	1.20	1.10	10.0
## 58	1	2012	SOUTH	2	TOTAL	S2TOTAL	1252	2.2	2.70	2.40	25.0
## 59	1	2012	SOUTH	2	TOTAL	S2TOTAL	1253	1.45	1.65	1.25	10.0
## 60	1	2012	SOUTH	2	TOTAL	S2TOTAL	1254	1.6	2.45	2.10	13.0
## 61	1	2012	SOUTH	2	TOTAL	S2TOTAL	1255	1.55	2.40	1.80	13.0
## 62	1	2012	SOUTH	2	TOTAL	S2TOTAL	1256	1.5	2.40	2.15	13.0
## 63	1	2012	SOUTH	2	TOTAL	S2TOTAL	1257	1.03	1.20	1.00	10.0
## 64	1	2012	SOUTH	2	TOTAL	S2TOTAL	1258	2.14	1.90	1.70	13.0
## 65	1	2012	SOUTH	2	TOTAL	S2TOTAL	1259	1.2	1.90	1.65	12.0

## 66	1	2012	SOUTH	2	TOTAL	S2TOTAL	1260	1.05	1.10	1.00	9.0
## 67	1	2012	SOUTH	2	TOTAL	S2TOTAL	2131	1.8	2.60	2.40	15.0
## 68	1	2012	SOUTH	2	TOTAL	S2TOTAL	2132	1.2	1.00	0.95	7.0
## 69	1	2012	SOUTH	2	TOTAL	S2TOTAL	2133	1.75	1.40	1.10	10.0
## 70	1	2012	SOUTH	2	TOTAL	S2TOTAL	2134	1.45	3.10	1.80	10.0
## 71	1	2012	SOUTH	2	TOTAL	S2TOTAL	2135	1.17	1.20	1.10	5.0
## 72	1	2012	SOUTH	2	TOTAL	S2TOTAL	2136	2.15	3.10	2.58	22.0
## 73	1	2012	SOUTH	2	TOTAL	S2TOTAL	2137	1.7	1.70	1.40	12.0
## 74	1	2012	SOUTH	2	TOTAL	S2TOTAL	3132	1.98	2.85	2.70	12.0
## 75	1	2012	SOUTH	2	TOTAL	S2TOTAL	3133	1.26	1.95	1.75	17.0
## 76	1	2012	SOUTH	2	TOTAL	S2TOTAL	3134	1.11	1.95	1.50	10.0
## 77	1	2012	SOUTH	2	TOTAL	S2TOTAL	3135	1.14	1.32	1.05	10.0
## 78	1	2012	SOUTH	2	TOTAL	S2TOTAL	3136	1.26	1.60	1.40	10.0
## 79	1	2012	SOUTH	2	TOTAL	S2TOTAL	3137	1.3	1.40	0.80	10.0
## 80	1	2012	SOUTH	2	TOTAL	S2TOTAL	3138	1.29	1.44	1.35	13.0
## 81	1	2012	SOUTH	2	TOTAL	S2TOTAL	3139	1.31	1.35	1.15	7.0
## 82	1	2012	SOUTH	2	TOTAL	S2TOTAL	3140	1.15	1.70	1.28	10.0
## 83	1	2012	SOUTH	2	TOTAL	S2TOTAL	3151	1.87	3.40	1.85	15.0
## 84	1	2012	SOUTH	2	TOTAL	S2TOTAL	3152	1.47	2.10	1.61	8.0
## 85	1	2012	SOUTH	2	TOTAL	S2TOTAL	3153	1.05	1.79	1.50	10.0
## 86	1	2012	SOUTH	2	TOTAL	S2TOTAL	3154	2.1	4.90	3.75	25.0
## 87	1	2012	SOUTH	2	TOTAL	S2TOTAL	3155	1.99	1.80	1.35	13.0
## 88	1	2012	SOUTH	2	TOTAL	S2TOTAL	3156	1.42	1.90	1.80	14.0
## 89	1	2012	SOUTH	2	TOTAL	S2TOTAL	3157	1.5	2.11	1.75	12.0
## 90	1	2012	SOUTH	2	TOTAL	S2TOTAL	3158	1.06	1.05	0.85	4.0
## 91	1	2012	SOUTH	2	TOTAL	S2TOTAL	3159	1.49	1.50	1.15	13.0
## 92	1	2012	SOUTH	2	TOTAL	S2TOTAL	3160	1.8	1.60	1.50	14.0
## 93	1	2012	SOUTH	2	TOTAL	S2TOTAL	3171	1.93	1.74	1.20	14.0
## 94	1	2012	SOUTH	2	TOTAL	S2TOTAL	3172	1.2	1.60	1.30	10.0
## 95	1	2012	SOUTH	2	TOTAL	S2TOTAL	3173	1.65	1.25	1.10	11.0
## 96	1	2012	SOUTH	2	TOTAL	S2TOTAL	3174	1.52	1.49	1.10	12.0
## 97	1	2012	SOUTH	2	TOTAL	S2TOTAL	3175	1.43	2.05	1.54	13.0
## 98	1	2012	SOUTH	2	TOTAL	S2TOTAL	3176	1.25	1.40	1.25	13.0
## 99	1	2012	SOUTH	2	TOTAL	S2TOTAL	3177	1.88	2.65	2.64	20.0
## 100	1	2012	SOUTH	2	TOTAL	S2TOTAL	3179	1.03	1.40	0.60	13.0
## 101	1	2012	SOUTH	2	TOTAL	S2TOTAL	3180	1.1	1.30	1.20	10.0
## 102	1	2012	SOUTH	2	TOTAL	S2TOTAL	3191	1.4	1.05	1.00	10.0
## 103	1	2012	SOUTH	2	TOTAL	S2TOTAL	3192	1.05	1.55	0.90	10.0
## 104	1	2012	SOUTH	2	TOTAL	S2TOTAL	3193	1.18	1.20	1.00	7.0
## 105	1	2012	SOUTH	2	TOTAL	S2TOTAL	3194	1.4	1.30	1.85	13.0
## 106	1	2012	SOUTH	2	TOTAL	S2TOTAL	3195	1.37	2.67	2.19	19.0
## 107	1	2012	SOUTH	2	TOTAL	S2TOTAL	3196	1.32	2.15	1.55	11.0
## 108	1	2012	SOUTH	2	MEGA	S2MEGA	182	1.55	2.20	1.20	20.0
## 109	1	2012	SOUTH	2	MEGA	S2MEGA	183	1.3	1.80	0.90	8.0
## 110	1	2012	SOUTH	2	MEGA	S2MEGA	184	1.24	1.20	1.20	25.0
## 111	1	2012	SOUTH	2	MEGA	S2MEGA	185	1.5	2.10	1.75	16.0
## 112	1	2012	SOUTH	2	MEGA	S2MEGA	186	1.65	2.50	2.20	15.0
## 113	1	2012	SOUTH	2	MEGA	S2MEGA	187	2.17	2.00	1.20	15.0
## 114	1	2012	SOUTH	2	MEGA	S2MEGA	188	1.28	1.60	1.50	10.0
## 115	1	2012	SOUTH	2	MEGA	S2MEGA	189	1.07	1.50	1.50	10.0
## 116	1	2012	SOUTH	2	MEGA	S2MEGA	190	0.67	1.00	0.80	8.0
## 117	1	2012	SOUTH	2	MEGA	S2MEGA	191	0.68	0.70	0.60	4.0
## 118	1	2012	SOUTH	2	MEGA	S2MEGA	192	1.87	1.60	1.40	9.0
## 119	1	2012	SOUTH	2	MEGA	S2MEGA	193	1.35	1.90	1.50	14.0

## 120	1	2012	SOUTH	2	MEGA	S2MEGA	194	1.75	2.10	2.10	15.0
## 121	1	2012	SOUTH	2	MESO	S2MESO	462	1.75	3.30	2.50	23.0
## 122	1	2012	SOUTH	2	MESO	S2MESO	463	1.64	2.30	2.00	14.0
## 123	1	2012	SOUTH	2	MESO	S2MESO	2138	1.42	0.90	0.80	10.0
## 124	1	2012	SOUTH	3	OPEN	S3OPEN	1301	dead	NA	NA	NA
## 125	1	2012	SOUTH	3	OPEN	S3OPEN	1302	0.9	1.30	1.10	11.0
## 126	1	2012	SOUTH	3	TOTAL	S3TOTAL	1061	dead	NA	NA	NA
## 127	1	2012	SOUTH	3	TOTAL	S3TOTAL	1062	1.8	2.60	2.60	15.0
## 128	1	2012	SOUTH	3	TOTAL	S3TOTAL	1063	2.47	3.10	2.20	18.0
## 129	1	2012	SOUTH	3	TOTAL	S3TOTAL	1064	2.15	1.60	1.10	17.0
## 130	1	2012	SOUTH	3	TOTAL	S3TOTAL	1066	1.7	2.50	2.15	15.0
## 131	1	2012	SOUTH	3	TOTAL	S3TOTAL	1066	1.9	1.80	1.50	20.0
## 132	1	2012	SOUTH	3	TOTAL	S3TOTAL	1067	1.95	2.10	1.90	13.0
## 133	1	2012	SOUTH	3	TOTAL	S3TOTAL	1068	1.8	1.70	1.40	13.0
## 134	1	2012	SOUTH	3	TOTAL	S3TOTAL	1069	1.4	2.00	1.60	14.0
## 135	1	2012	SOUTH	3	TOTAL	S3TOTAL	1070	1	1.30	1.20	7.0
## 136	1	2012	SOUTH	3	TOTAL	S3TOTAL	2139	1.75	1.20	1.10	13.0
## 137	1	2012	SOUTH	3	TOTAL	S3TOTAL	2140	1.28	1.50	0.95	4.0
## 138	1	2012	SOUTH	3	TOTAL	S3TOTAL	2151	1	1.40	1.20	4.0
## 139	1	2012	SOUTH	3	TOTAL	S3TOTAL	2152	1.45	1.50	1.30	10.0
## 140	1	2012	SOUTH	3	TOTAL	S3TOTAL	2153	1	1.00	0.75	8.0
## 141	1	2012	SOUTH	3	TOTAL	S3TOTAL	2154	1.03	1.00	0.90	6.0
## 142	1	2012	SOUTH	3	TOTAL	S3TOTAL	2155	1.51	2.00	1.80	12.0
## 143	1	2012	SOUTH	3	TOTAL	S3TOTAL	2156	1.17	1.10	0.90	10.0
## 144	1	2012	SOUTH	3	TOTAL	S3TOTAL	2157	1.33	1.90	1.85	14.0
## 145	1	2012	SOUTH	3	TOTAL	S3TOTAL	2158	1.3	1.10	0.85	8.0
## 146	1	2012	SOUTH	3	TOTAL	S3TOTAL	2159	1.13	1.10	0.90	10.0
## 147	1	2012	SOUTH	3	TOTAL	S3TOTAL	2160	1.58	1.40	1.40	13.0
## 148	1	2012	SOUTH	3	TOTAL	S3TOTAL	2171	1.06	1.40	1.00	5.0
## 149	1	2012	SOUTH	3	TOTAL	S3TOTAL	2172	1.05	1.40	0.95	7.0
## 150	1	2012	SOUTH	3	TOTAL	S3TOTAL	2173	1.45	1.60	1.10	6.0
## 151	1	2012	SOUTH	3	TOTAL	S3TOTAL	2174	1.15	1.10	0.90	5.0
## 152	1	2012	SOUTH	3	TOTAL	S3TOTAL	2175	1.42	1.45	1.30	13.0
## 153	1	2012	SOUTH	3	TOTAL	S3TOTAL	2176	1.02	1.20	1.00	8.0
## 154	1	2012	SOUTH	3	TOTAL	S3TOTAL	2177	1.4	1.20	1.00	9.0
## 155	1	2012	SOUTH	3	TOTAL	S3TOTAL	2178	1.45	2.10	2.05	15.0
## 156	1	2012	SOUTH	3	MESO	S3MESO	1421	1.95	2.20	1.60	13.0
## 157	1	2012	SOUTH	3	MESO	S3MESO	1422	dead	NA	NA	NA
##	FLOWERS	BUDS	FRUITS	ANT							
## 1	0	0	10	CS							
## 2	0	0	150	TP							
## 3	2	1	50	TP							
## 4	0	0	75	CS							
## 5	0	0	20	CS							
## 6	0	0	0	E							
## 7	0	0	0	CS							
## 8	0	0	25	CS							
## 9	0	0	0	TP							
## 10	0	0	50	TP							
## 11	0	0	5	CS							
## 12	0	0	60	TP							
## 13	0	0	60	TP							
## 14	2	0	60	CS							
## 15	2	0	0	CS							

## 16	0	0	0	TP
## 17	0	0	0	TP
## 18	0	0	0	CS
## 19	0	0	0	CM
## 20	0	0	0	TP
## 21	NA	NA	NA	
## 22	0	0	5	CS
## 23	0	0	45	CS
## 24	40	50	35	CS
## 25	8	2	65	CS
## 26	0	0	20	TP
## 27	0	0	70	CS
## 28	0	0	125	CM
## 29	0	0	200	CM
## 30	0	0	10	CS
## 31	0	0	0	CS
## 32	0	0	35	TP
## 33	0	0	300	CM
## 34	2	2	100	CS
## 35	0	0	30	CM
## 36	0	0	50	TP
## 37	0	0	10	CM
## 38	0	0	25	CS
## 39	0	0	15	TP
## 40	0	0	0	TP
## 41	0	0	15	TP
## 42	0	0	0	TP
## 43	0	0	40	TP
## 44	0	0	0	TP
## 45	0	0	15	CM
## 46	0	0	0	CM
## 47	0	0	0	TP
## 48	0	0	0	TP
## 49	0	0	1	TP
## 50	0	0	20	TP
## 51	0	0	0	TP
## 52	0	0	0	TP
## 53	0	0	20	TP
## 54	0	0	0	TP
## 55	0	0	0	CN
## 56	0	0	0	CN
## 57	0	0	0	TP
## 58	0	0	5	TP
## 59	0	0	0	TP
## 60	0	0	25	TP
## 61	0	0	25	TP
## 62	0	0	20	TP
## 63	0	0	0	TP
## 64	0	0	10	CS
## 65	1	0	25	CS
## 66	0	0	0	TP
## 67	0	0	10	TP
## 68	0	0	0	TP
## 69	0	0	0	TP

## 70	0	0	0	TP
## 71	0	0	0	TP
## 72	0	0	0	CS
## 73	0	0	0	CS
## 74	0	0	25	AB_TP
## 75	0	0	0	TP
## 76	0	0	0	TP
## 77	0	0	0	TP
## 78	0	0	0	CS
## 79	0	0	0	CS
## 80	0	0	0	CS
## 81	0	0	0	CS
## 82	0	0	5	CS
## 83	6	0	0	CS
## 84	0	0	0	CS
## 85	0	0	1	CS
## 86	0	0	25	CS
## 87	0	0	0	CS
## 88	0	0	0	CS
## 89	0	0	10	CS
## 90	0	0	0	CS
## 91	0	0	35	CS
## 92	0	0	0	CS
## 93	0	0	0	CS
## 94	0	0	0	CS
## 95	0	0	0	CS
## 96	0	0	20	CS
## 97	0	0	0	CS
## 98	0	0	0	CM
## 99	0	0	100	CM
## 100	0	0	0	CS
## 101	0	0	0	CS
## 102	0	0	0	CS
## 103	0	0	0	CM
## 104	0	0	0	TP
## 105	0	0	30	CS
## 106	0	0	50	TP
## 107	0	0	10	CS
## 108	0	0	0	CS
## 109	0	0	15	CS
## 110	0	0	10	CS
## 111	5	0	200	CS
## 112	0	0	80	CS
## 113	0	0	150	TP
## 114	0	0	40	TP
## 115	0	0	60	TP
## 116	0	0	0	CS
## 117	0	0	0	TP
## 118	0	0	40	CS
## 119	0	0	20	CS
## 120	0	0	75	TP
## 121	0	0	20	CM
## 122	0	0	0	TP
## 123	0	0	0	E

## 124	NA	NA	NA	
## 125	0	0	0	TP
## 126	NA	NA	NA	
## 127	0	0	50	TP
## 128	0	0	0	TP
## 129	0	0	0	TP
## 130	0	0	2	TP
## 131	0	0	25	TP
## 132	0	0	0	TP
## 133	0	0	0	TP
## 134	0	0	0	TP
## 135	0	0	0	TP
## 136	0	0	0	TP
## 137	0	0	0	TP
## 138	0	0	0	TP
## 139	0	0	0	TP
## 140	0	0	0	TP
## 141	0	0	0	TP
## 142	0	0	0	TP
## 143	0	0	0	TP
## 144	0	0	0	TP
## 145	0	0	0	TP
## 146	0	0	0	TP
## 147	0	0	0	TP
## 148	0	0	8	TP
## 149	0	0	0	TP
## 150	0	0	0	TP
## 151	0	0	0	TP
## 152	0	0	0	TP
## 153	0	0	0	TP
## 154	0	0	0	TP
## 155	0	0	20	TP
## 156	0	0	2	CS
## 157	NA	NA	NA	

3. Explore the data set

```
head(acacia)
```

##	SURVEY	YEAR	SITE	BLOCK	TREATMENT	PLOT	ID	HEIGHT	AXIS1	AXIS2	CIRC
## 1	1	2012	SOUTH	1	TOTAL	S1TOTAL	581	2.25	2.75	2.15	20
## 2	1	2012	SOUTH	1	TOTAL	S1TOTAL	582	2.65	4.10	3.90	28
## 3	1	2012	SOUTH	1	TOTAL	S1TOTAL	3111	1.5	1.70	0.85	17
## 4	1	2012	SOUTH	1	TOTAL	S1TOTAL	3112	2.01	1.80	1.60	12
## 5	1	2012	SOUTH	1	TOTAL	S1TOTAL	3113	1.75	1.84	1.42	13
## 6	1	2012	SOUTH	1	TOTAL	S1TOTAL	3114	1.65	1.62	0.85	15
##	FLOWERS	BUDS	FRUITS	ANT							
## 1	0	0	10	CS							
## 2	0	0	150	TP							
## 3	2	1	50	TP							
## 4	0	0	75	CS							
## 5	0	0	20	CS							
## 6	0	0	0	E							

```
summary(acacia)
```

```
##      SURVEY      YEAR      SITE      BLOCK
## Min.   :1   Min.   :2012   Length:157   Min.   :1.000
## 1st Qu.:1   1st Qu.:2012   Class :character 1st Qu.:2.000
## Median :1   Median :2012   Mode  :character Median :2.000
## Mean   :1   Mean   :2012                   Mean   :2.089
## 3rd Qu.:1   3rd Qu.:2012                   3rd Qu.:2.000
## Max.   :1   Max.   :2012                   Max.   :3.000
##
##      TREATMENT      PLOT      ID      HEIGHT
## Length:157      Length:157   Min.   : 101   Length:157
## Class :character Class :character 1st Qu.:1062   Class :character
## Mode  :character Mode  :character Median :1301   Mode  :character
##                                     Mean   :1743
##                                     3rd Qu.:3118
##                                     Max.   :3199
##
##      AXIS1      AXIS2      CIRC      FLOWERS
## Min.   :0.700   Min.   :0.550   Min.   : 4.00   Min.   : 0.0000
## 1st Qu.:1.400   1st Qu.:1.100   1st Qu.:10.00   1st Qu.: 0.0000
## Median :1.800   Median :1.490   Median :13.00   Median : 0.0000
## Mean   :1.972   Mean   :1.636   Mean   :13.76   Mean   : 0.4444
## 3rd Qu.:2.350   3rd Qu.:2.000   3rd Qu.:16.00   3rd Qu.: 0.0000
## Max.   :5.550   Max.   :4.820   Max.   :35.20   Max.   :40.0000
## NA's   :4      NA's   :4      NA's   :4      NA's   :4
##
##      BUDS      FRUITS      ANT
## Min.   : 0.0000   Min.   : 0.00   Length:157
## 1st Qu.: 0.0000   1st Qu.: 0.00   Class :character
## Median : 0.0000   Median : 0.00   Mode  :character
## Mean   : 0.3595   Mean   : 20.03
## 3rd Qu.: 0.0000   3rd Qu.: 25.00
## Max.   :50.0000   Max.   :300.00
## NA's   :4      NA's   :4
```

```
colnames(acacia)
```

```
## [1] "SURVEY" "YEAR" "SITE" "BLOCK" "TREATMENT" "PLOT"
## [7] "ID" "HEIGHT" "AXIS1" "AXIS2" "CIRC" "FLOWERS"
## [13] "BUDS" "FRUITS" "ANT"
```

```
rownames(acacia)
```

```
## [1] "1" "2" "3" "4" "5" "6" "7" "8" "9" "10" "11" "12"
## [13] "13" "14" "15" "16" "17" "18" "19" "20" "21" "22" "23" "24"
## [25] "25" "26" "27" "28" "29" "30" "31" "32" "33" "34" "35" "36"
## [37] "37" "38" "39" "40" "41" "42" "43" "44" "45" "46" "47" "48"
## [49] "49" "50" "51" "52" "53" "54" "55" "56" "57" "58" "59" "60"
## [61] "61" "62" "63" "64" "65" "66" "67" "68" "69" "70" "71" "72"
## [73] "73" "74" "75" "76" "77" "78" "79" "80" "81" "82" "83" "84"
## [85] "85" "86" "87" "88" "89" "90" "91" "92" "93" "94" "95" "96"
## [97] "97" "98" "99" "100" "101" "102" "103" "104" "105" "106" "107" "108"
## [109] "109" "110" "111" "112" "113" "114" "115" "116" "117" "118" "119" "120"
## [121] "121" "122" "123" "124" "125" "126" "127" "128" "129" "130" "131" "132"
## [133] "133" "134" "135" "136" "137" "138" "139" "140" "141" "142" "143" "144"
```



```
## [145] "145" "146" "147" "148" "149" "150" "151" "152" "153" "154" "155" "156"
## [157] "157"
```

```
class(acacia$SURVEY)
```

```
## [1] "integer"
```

```
# the sapply function allows to apply a function to a list of objects
# a data frame is a list of vectors of the same length
sapply(acacia, class)
```

```
##      SURVEY      YEAR      SITE      BLOCK  TREATMENT      PLOT
## "integer" "integer" "character" "integer" "character" "character"
##      ID      HEIGHT      AXIS1      AXIS2      CIRC      FLOWERS
## "integer" "character" "numeric" "numeric" "numeric" "integer"
##      BUDS      FRUITS      ANT
## "integer" "integer" "character"
```

```
View(acacia)
```

Make sure that everything that is a number is actually `typeof` numeric or double or integer.

One way to do this is with the function `summary`, and checking at the type of data on each column visually.

Another way is using the `typeof()` function:

```
typeof(acacia[, "HEIGHT"])
```

```
## [1] "character"
```

```
acacia$HEIGHT
```

```
## [1] "2.25" "2.65" "1.5" "2.01" "1.75" "1.65" "1.2" "1.45" "1.87" "2.38"
## [11] "2.58" "2.65" "2.35" "1.88" "2.32" "2.39" "2.2" "1.05" "2" "1.28"
## [21] "dead" "1.4" "1.9" "1.75" "1.8" "2.7" "2.02" "1.9" "1.85" "1.65"
## [31] "1.4" "2.5" "2.05" "2.26" "2.13" "1.8" "1.85" "1.5" "1.87" "1.58"
## [41] "2.05" "1.75" "1.49" "1.28" "1.49" "1.07" "1.48" "1.25" "1.41" "1.6"
## [51] "1.2" "1.49" "1.5" "1.65" "1.13" "1.25" "1.1" "2.2" "1.45" "1.6"
## [61] "1.55" "1.5" "1.03" "2.14" "1.2" "1.05" "1.8" "1.2" "1.75" "1.45"
## [71] "1.17" "2.15" "1.7" "1.98" "1.26" "1.11" "1.14" "1.26" "1.3" "1.29"
## [81] "1.31" "1.15" "1.87" "1.47" "1.05" "2.1" "1.99" "1.42" "1.5" "1.06"
## [91] "1.49" "1.8" "1.93" "1.2" "1.65" "1.52" "1.43" "1.25" "1.88" "1.03"
## [101] "1.1" "1.4" "1.05" "1.18" "1.4" "1.37" "1.32" "1.55" "1.3" "1.24"
## [111] "1.5" "1.65" "2.17" "1.28" "1.07" "0.67" "0.68" "1.87" "1.35" "1.75"
## [121] "1.75" "1.64" "1.42" "dead" "0.9" "dead" "1.8" "2.47" "2.15" "1.7"
## [131] "1.9" "1.95" "1.8" "1.4" "1" "1.75" "1.28" "1" "1.45" "1"
## [141] "1.03" "1.51" "1.17" "1.33" "1.3" "1.13" "1.58" "1.06" "1.05" "1.45"
## [151] "1.15" "1.42" "1.02" "1.4" "1.45" "1.95" "dead"
```

We identified a column that has problematic data. We need to fix it!

Cleaning our raw data

3.2 Assigning NA to missing/incorrect values

We are going to read the data table again, but we are gonna assign NA to the “dead” value that we do not want in our “HEIGHT” column.

```
acacia <- read.csv(file = "/Users/lunasare/Desktop/data-science-course/fall-2022/data-raw/ACACIA_DREPAN",
  sep = "\t",
  na.strings = "dead")
```

Let's check if this worked:

```
acacia$HEIGHT

## [1] 2.25 2.65 1.50 2.01 1.75 1.65 1.20 1.45 1.87 2.38 2.58 2.65 2.35 1.88 2.32
## [16] 2.39 2.20 1.05 2.00 1.28 NA 1.40 1.90 1.75 1.80 2.70 2.02 1.90 1.85 1.65
## [31] 1.40 2.50 2.05 2.26 2.13 1.80 1.85 1.50 1.87 1.58 2.05 1.75 1.49 1.28 1.49
## [46] 1.07 1.48 1.25 1.41 1.60 1.20 1.49 1.50 1.65 1.13 1.25 1.10 2.20 1.45 1.60
## [61] 1.55 1.50 1.03 2.14 1.20 1.05 1.80 1.20 1.75 1.45 1.17 2.15 1.70 1.98 1.26
## [76] 1.11 1.14 1.26 1.30 1.29 1.31 1.15 1.87 1.47 1.05 2.10 1.99 1.42 1.50 1.06
## [91] 1.49 1.80 1.93 1.20 1.65 1.52 1.43 1.25 1.88 1.03 1.10 1.40 1.05 1.18 1.40
## [106] 1.37 1.32 1.55 1.30 1.24 1.50 1.65 2.17 1.28 1.07 0.67 0.68 1.87 1.35 1.75
## [121] 1.75 1.64 1.42 NA 0.90 NA 1.80 2.47 2.15 1.70 1.90 1.95 1.80 1.40 1.00
## [136] 1.75 1.28 1.00 1.45 1.00 1.03 1.51 1.17 1.33 1.30 1.13 1.58 1.06 1.05 1.45
## [151] 1.15 1.42 1.02 1.40 1.45 1.95 NA

typeof(acacia$HEIGHT)

## [1] "double"

class(acacia$HEIGHT)

## [1] "numeric"
```

4. Visualize our data

For this, we are using the `ggplot2` package. Let's install it and load it:

```
# install.packages("ggplot2")
library(ggplot2)
```

Now we are gonna create our first plotting layer with the function `ggplot`.

```
colnames(acacia)

## [1] "SURVEY" "YEAR" "SITE" "BLOCK" "TREATMENT" "PLOT"
## [7] "ID" "HEIGHT" "AXIS1" "AXIS2" "CIRC" "FLOWERS"
## [13] "BUDS" "FRUITS" "ANT"

acacia$CIRC

## [1] 20.0 28.0 17.0 12.0 13.0 15.0 9.0 12.2 13.0 35.0 24.0 27.0 20.0 28.0 30.0
## [16] 13.0 10.0 8.0 10.0 10.0 NA 18.0 15.0 16.0 16.0 35.2 17.0 19.0 19.0 17.0
## [31] 14.0 22.0 33.0 33.0 20.0 22.0 20.0 15.0 13.0 11.0 17.0 16.0 13.0 10.0 13.0
## [46] 11.0 9.0 10.0 14.0 13.0 14.0 8.0 14.0 20.0 10.0 10.0 10.0 25.0 10.0 13.0
## [61] 13.0 13.0 10.0 13.0 12.0 9.0 15.0 7.0 10.0 10.0 5.0 22.0 12.0 12.0 17.0
## [76] 10.0 10.0 10.0 10.0 13.0 7.0 10.0 15.0 8.0 10.0 25.0 13.0 14.0 12.0 4.0
## [91] 13.0 14.0 14.0 10.0 11.0 12.0 13.0 13.0 20.0 13.0 10.0 10.0 10.0 7.0 13.0
## [106] 19.0 11.0 20.0 8.0 25.0 16.0 15.0 15.0 10.0 10.0 8.0 4.0 9.0 14.0 15.0
## [121] 23.0 14.0 10.0 NA 11.0 NA 15.0 18.0 17.0 15.0 20.0 13.0 13.0 14.0 7.0
## [136] 13.0 4.0 4.0 10.0 8.0 6.0 12.0 10.0 14.0 8.0 10.0 13.0 5.0 7.0 6.0
## [151] 5.0 13.0 8.0 9.0 15.0 13.0 NA
```

The `ggplot` function creates a blank canvas. The canvas contains our data, and the variables that we want to plot.

A scatter plot can be created with the function `geom_point()`

colors()

## [1]	"white"	"aliceblue"	"antiquewhite"
## [4]	"antiquewhite1"	"antiquewhite2"	"antiquewhite3"
## [7]	"antiquewhite4"	"aquamarine"	"aquamarine1"
## [10]	"aquamarine2"	"aquamarine3"	"aquamarine4"
## [13]	"azure"	"azure1"	"azure2"
## [16]	"azure3"	"azure4"	"beige"
## [19]	"bisque"	"bisque1"	"bisque2"
## [22]	"bisque3"	"bisque4"	"black"
## [25]	"blanchedalmond"	"blue"	"blue1"
## [28]	"blue2"	"blue3"	"blue4"
## [31]	"blueviolet"	"brown"	"brown1"
## [34]	"brown2"	"brown3"	"brown4"
## [37]	"burlywood"	"burlywood1"	"burlywood2"
## [40]	"burlywood3"	"burlywood4"	"cadetblue"
## [43]	"cadetblue1"	"cadetblue2"	"cadetblue3"
## [46]	"cadetblue4"	"chartreuse"	"chartreuse1"
## [49]	"chartreuse2"	"chartreuse3"	"chartreuse4"
## [52]	"chocolate"	"chocolate1"	"chocolate2"
## [55]	"chocolate3"	"chocolate4"	"coral"
## [58]	"coral1"	"coral2"	"coral3"
## [61]	"coral4"	"cornflowerblue"	"cornsilk"
## [64]	"cornsilk1"	"cornsilk2"	"cornsilk3"
## [67]	"cornsilk4"	"cyan"	"cyan1"
## [70]	"cyan2"	"cyan3"	"cyan4"
## [73]	"darkblue"	"darkcyan"	"darkgoldenrod"
## [76]	"darkgoldenrod1"	"darkgoldenrod2"	"darkgoldenrod3"
## [79]	"darkgoldenrod4"	"darkgray"	"darkgreen"
## [82]	"darkgrey"	"darkkhaki"	"darkmagenta"
## [85]	"darkolivegreen"	"darkolivegreen1"	"darkolivegreen2"
## [88]	"darkolivegreen3"	"darkolivegreen4"	"darkorange"
## [91]	"darkorange1"	"darkorange2"	"darkorange3"
## [94]	"darkorange4"	"darkorchid"	"darkorchid1"
## [97]	"darkorchid2"	"darkorchid3"	"darkorchid4"
## [100]	"darkred"	"darksalmon"	"darkseagreen"
## [103]	"darkseagreen1"	"darkseagreen2"	"darkseagreen3"
## [106]	"darkseagreen4"	"darkslateblue"	"darkslategray"
## [109]	"darkslategray1"	"darkslategray2"	"darkslategray3"
## [112]	"darkslategray4"	"darkslategrey"	"darkturquoise"
## [115]	"darkviolet"	"deeppink"	"deeppink1"
## [118]	"deeppink2"	"deeppink3"	"deeppink4"
## [121]	"deepskyblue"	"deepskyblue1"	"deepskyblue2"
## [124]	"deepskyblue3"	"deepskyblue4"	"dimgray"
## [127]	"dimgrey"	"dodgerblue"	"dodgerblue1"
## [130]	"dodgerblue2"	"dodgerblue3"	"dodgerblue4"
## [133]	"firebrick"	"firebrick1"	"firebrick2"
## [136]	"firebrick3"	"firebrick4"	"floralwhite"
## [139]	"forestgreen"	"gainsboro"	"ghostwhite"
## [142]	"gold"	"gold1"	"gold2"
## [145]	"gold3"	"gold4"	"goldenrod"
## [148]	"goldenrod1"	"goldenrod2"	"goldenrod3"
## [151]	"goldenrod4"	"gray"	"gray0"
## [154]	"gray1"	"gray2"	"gray3"

## [157]	"gray4"	"gray5"	"gray6"
## [160]	"gray7"	"gray8"	"gray9"
## [163]	"gray10"	"gray11"	"gray12"
## [166]	"gray13"	"gray14"	"gray15"
## [169]	"gray16"	"gray17"	"gray18"
## [172]	"gray19"	"gray20"	"gray21"
## [175]	"gray22"	"gray23"	"gray24"
## [178]	"gray25"	"gray26"	"gray27"
## [181]	"gray28"	"gray29"	"gray30"
## [184]	"gray31"	"gray32"	"gray33"
## [187]	"gray34"	"gray35"	"gray36"
## [190]	"gray37"	"gray38"	"gray39"
## [193]	"gray40"	"gray41"	"gray42"
## [196]	"gray43"	"gray44"	"gray45"
## [199]	"gray46"	"gray47"	"gray48"
## [202]	"gray49"	"gray50"	"gray51"
## [205]	"gray52"	"gray53"	"gray54"
## [208]	"gray55"	"gray56"	"gray57"
## [211]	"gray58"	"gray59"	"gray60"
## [214]	"gray61"	"gray62"	"gray63"
## [217]	"gray64"	"gray65"	"gray66"
## [220]	"gray67"	"gray68"	"gray69"
## [223]	"gray70"	"gray71"	"gray72"
## [226]	"gray73"	"gray74"	"gray75"
## [229]	"gray76"	"gray77"	"gray78"
## [232]	"gray79"	"gray80"	"gray81"
## [235]	"gray82"	"gray83"	"gray84"
## [238]	"gray85"	"gray86"	"gray87"
## [241]	"gray88"	"gray89"	"gray90"
## [244]	"gray91"	"gray92"	"gray93"
## [247]	"gray94"	"gray95"	"gray96"
## [250]	"gray97"	"gray98"	"gray99"
## [253]	"gray100"	"green"	"green1"
## [256]	"green2"	"green3"	"green4"
## [259]	"greenyellow"	"grey"	"grey0"
## [262]	"grey1"	"grey2"	"grey3"
## [265]	"grey4"	"grey5"	"grey6"
## [268]	"grey7"	"grey8"	"grey9"
## [271]	"grey10"	"grey11"	"grey12"
## [274]	"grey13"	"grey14"	"grey15"
## [277]	"grey16"	"grey17"	"grey18"
## [280]	"grey19"	"grey20"	"grey21"
## [283]	"grey22"	"grey23"	"grey24"
## [286]	"grey25"	"grey26"	"grey27"
## [289]	"grey28"	"grey29"	"grey30"
## [292]	"grey31"	"grey32"	"grey33"
## [295]	"grey34"	"grey35"	"grey36"
## [298]	"grey37"	"grey38"	"grey39"
## [301]	"grey40"	"grey41"	"grey42"
## [304]	"grey43"	"grey44"	"grey45"
## [307]	"grey46"	"grey47"	"grey48"
## [310]	"grey49"	"grey50"	"grey51"
## [313]	"grey52"	"grey53"	"grey54"
## [316]	"grey55"	"grey56"	"grey57"

## [319]	"grey58"	"grey59"	"grey60"
## [322]	"grey61"	"grey62"	"grey63"
## [325]	"grey64"	"grey65"	"grey66"
## [328]	"grey67"	"grey68"	"grey69"
## [331]	"grey70"	"grey71"	"grey72"
## [334]	"grey73"	"grey74"	"grey75"
## [337]	"grey76"	"grey77"	"grey78"
## [340]	"grey79"	"grey80"	"grey81"
## [343]	"grey82"	"grey83"	"grey84"
## [346]	"grey85"	"grey86"	"grey87"
## [349]	"grey88"	"grey89"	"grey90"
## [352]	"grey91"	"grey92"	"grey93"
## [355]	"grey94"	"grey95"	"grey96"
## [358]	"grey97"	"grey98"	"grey99"
## [361]	"grey100"	"honeydew"	"honeydew1"
## [364]	"honeydew2"	"honeydew3"	"honeydew4"
## [367]	"hotpink"	"hotpink1"	"hotpink2"
## [370]	"hotpink3"	"hotpink4"	"indianred"
## [373]	"indianred1"	"indianred2"	"indianred3"
## [376]	"indianred4"	"ivory"	"ivory1"
## [379]	"ivory2"	"ivory3"	"ivory4"
## [382]	"khaki"	"khaki1"	"khaki2"
## [385]	"khaki3"	"khaki4"	"lavender"
## [388]	"lavenderblush"	"lavenderblush1"	"lavenderblush2"
## [391]	"lavenderblush3"	"lavenderblush4"	"lawngreen"
## [394]	"lemonchiffon"	"lemonchiffon1"	"lemonchiffon2"
## [397]	"lemonchiffon3"	"lemonchiffon4"	"lightblue"
## [400]	"lightblue1"	"lightblue2"	"lightblue3"
## [403]	"lightblue4"	"lightcoral"	"lightcyan"
## [406]	"lightcyan1"	"lightcyan2"	"lightcyan3"
## [409]	"lightcyan4"	"lightgoldenrod"	"lightgoldenrod1"
## [412]	"lightgoldenrod2"	"lightgoldenrod3"	"lightgoldenrod4"
## [415]	"lightgoldenrodyellow"	"lightgray"	"lightgreen"
## [418]	"lightgrey"	"lightpink"	"lightpink1"
## [421]	"lightpink2"	"lightpink3"	"lightpink4"
## [424]	"lightsalmon"	"lightsalmon1"	"lightsalmon2"
## [427]	"lightsalmon3"	"lightsalmon4"	"lightseagreen"
## [430]	"lightskyblue"	"lightskyblue1"	"lightskyblue2"
## [433]	"lightskyblue3"	"lightskyblue4"	"lightslateblue"
## [436]	"lightslategray"	"lightslategrey"	"lightsteelblue"
## [439]	"lightsteelblue1"	"lightsteelblue2"	"lightsteelblue3"
## [442]	"lightsteelblue4"	"lightyellow"	"lightyellow1"
## [445]	"lightyellow2"	"lightyellow3"	"lightyellow4"
## [448]	"limegreen"	"linen"	"magenta"
## [451]	"magenta1"	"magenta2"	"magenta3"
## [454]	"magenta4"	"maroon"	"maroon1"
## [457]	"maroon2"	"maroon3"	"maroon4"
## [460]	"mediumaquamarine"	"mediumblue"	"mediumorchid"
## [463]	"mediumorchid1"	"mediumorchid2"	"mediumorchid3"
## [466]	"mediumorchid4"	"mediumpurple"	"mediumpurple1"
## [469]	"mediumpurple2"	"mediumpurple3"	"mediumpurple4"
## [472]	"mediumseagreen"	"mediumslateblue"	"mediumspringgreen"
## [475]	"mediumturquoise"	"mediumvioletred"	"midnightblue"
## [478]	"mintcream"	"mistyrose"	"mistyrose1"

## [481]	"mistyrose2"	"mistyrose3"	"mistyrose4"
## [484]	"moccasin"	"navajowhite"	"navajowhite1"
## [487]	"navajowhite2"	"navajowhite3"	"navajowhite4"
## [490]	"navy"	"navyblue"	"oldlace"
## [493]	"olivedrab"	"olivedrab1"	"olivedrab2"
## [496]	"olivedrab3"	"olivedrab4"	"orange"
## [499]	"orange1"	"orange2"	"orange3"
## [502]	"orange4"	"orangered"	"orangered1"
## [505]	"orangered2"	"orangered3"	"orangered4"
## [508]	"orchid"	"orchid1"	"orchid2"
## [511]	"orchid3"	"orchid4"	"palegoldenrod"
## [514]	"palegreen"	"palegreen1"	"palegreen2"
## [517]	"palegreen3"	"palegreen4"	"paleturquoise"
## [520]	"paleturquoise1"	"paleturquoise2"	"paleturquoise3"
## [523]	"paleturquoise4"	"palevioletred"	"palevioletred1"
## [526]	"palevioletred2"	"palevioletred3"	"palevioletred4"
## [529]	"papayawhip"	"peachpuff"	"peachpuff1"
## [532]	"peachpuff2"	"peachpuff3"	"peachpuff4"
## [535]	"peru"	"pink"	"pink1"
## [538]	"pink2"	"pink3"	"pink4"
## [541]	"plum"	"plum1"	"plum2"
## [544]	"plum3"	"plum4"	"powderblue"
## [547]	"purple"	"purple1"	"purple2"
## [550]	"purple3"	"purple4"	"red"
## [553]	"red1"	"red2"	"red3"
## [556]	"red4"	"rosybrown"	"rosybrown1"
## [559]	"rosybrown2"	"rosybrown3"	"rosybrown4"
## [562]	"royalblue"	"royalblue1"	"royalblue2"
## [565]	"royalblue3"	"royalblue4"	"saddlebrown"
## [568]	"salmon"	"salmon1"	"salmon2"
## [571]	"salmon3"	"salmon4"	"sandybrown"
## [574]	"seagreen"	"seagreen1"	"seagreen2"
## [577]	"seagreen3"	"seagreen4"	"seashell"
## [580]	"seashell1"	"seashell2"	"seashell3"
## [583]	"seashell4"	"sienna"	"sienna1"
## [586]	"sienna2"	"sienna3"	"sienna4"
## [589]	"skyblue"	"skyblue1"	"skyblue2"
## [592]	"skyblue3"	"skyblue4"	"slateblue"
## [595]	"slateblue1"	"slateblue2"	"slateblue3"
## [598]	"slateblue4"	"slategray"	"slategray1"
## [601]	"slategray2"	"slategray3"	"slategray4"
## [604]	"slategrey"	"snow"	"snow1"
## [607]	"snow2"	"snow3"	"snow4"
## [610]	"springgreen"	"springgreen1"	"springgreen2"
## [613]	"springgreen3"	"springgreen4"	"steelblue"
## [616]	"steelblue1"	"steelblue2"	"steelblue3"
## [619]	"steelblue4"	"tan"	"tan1"
## [622]	"tan2"	"tan3"	"tan4"
## [625]	"thistle"	"thistle1"	"thistle2"
## [628]	"thistle3"	"thistle4"	"tomato"
## [631]	"tomato1"	"tomato2"	"tomato3"
## [634]	"tomato4"	"turquoise"	"turquoise1"
## [637]	"turquoise2"	"turquoise3"	"turquoise4"
## [640]	"violet"	"violetred"	"violetred1"

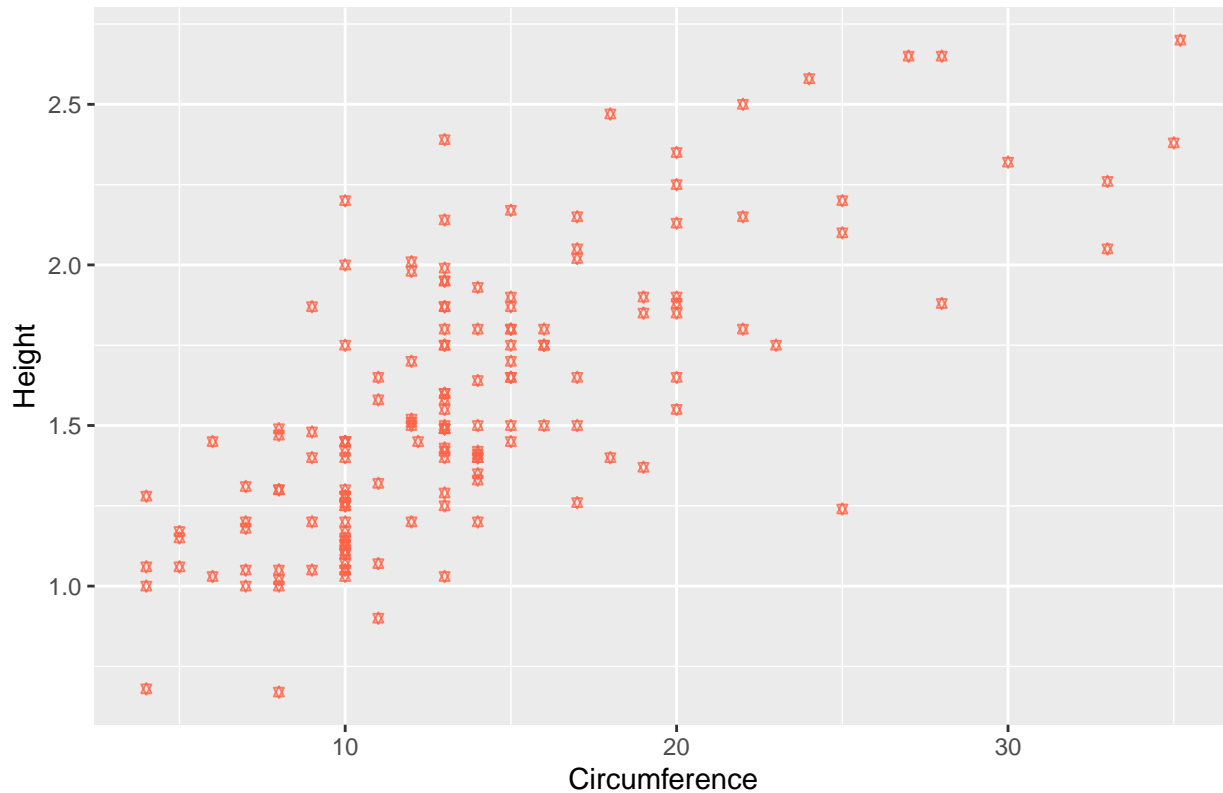
```
## [643] "violetred2"      "violetred3"      "violetred4"
## [646] "wheat"           "wheat1"          "wheat2"
## [649] "wheat3"          "wheat4"          "whitesmoke"
## [652] "yellow"          "yellow1"         "yellow2"
## [655] "yellow3"         "yellow4"         "yellowgreen"
```

```
?pch
```

```
ggplot(data = acacia, mapping = aes(x = CIRC, y = HEIGHT)) +
  geom_point(size = 1, color = "tomato", alpha = 0.8, shape = 11) +
  labs(x = "Circumference", y = "Height", title = "Data from UHURU Acacia survey")
```

```
## Warning: Removed 4 rows containing missing values (geom_point).
```

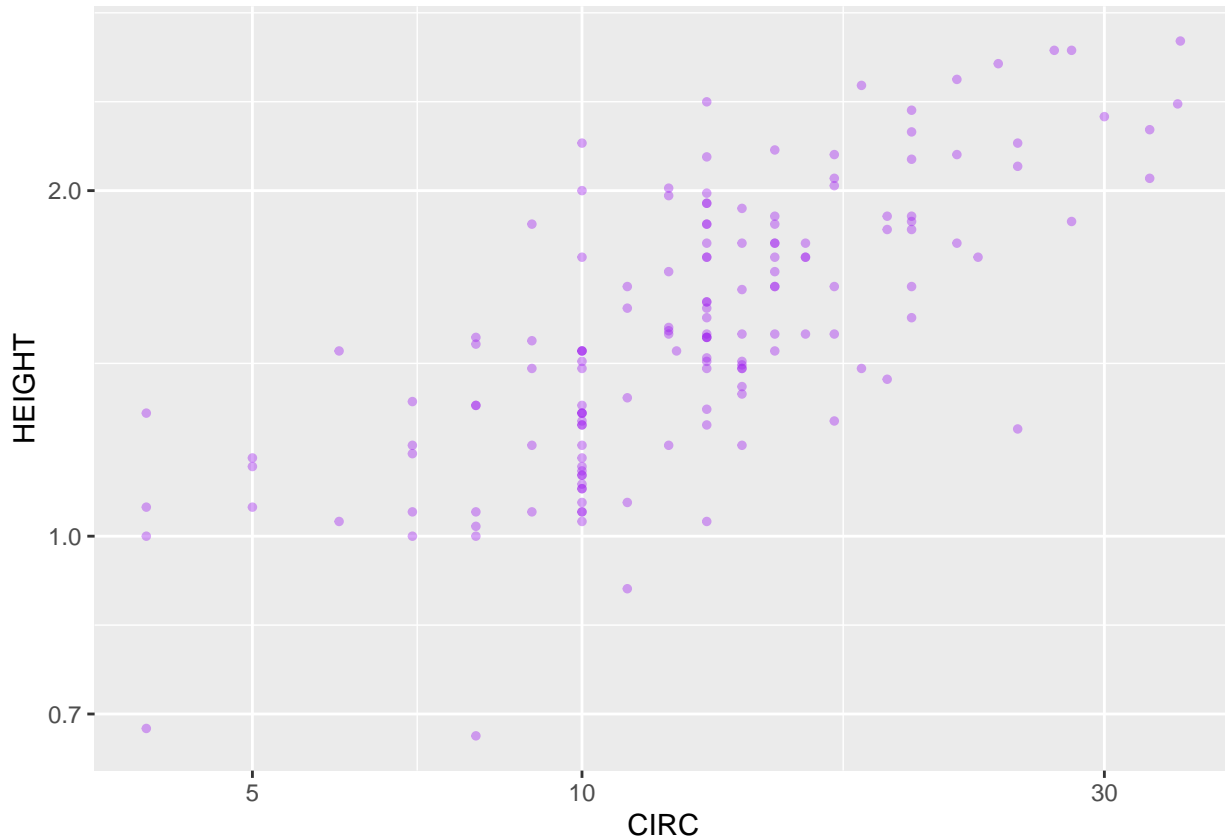
Data from UHURU Acacia survey



To rescale the plotting of the axis to log scale we use the function `scale_y_log10()` and `scale_x_log10()`:

```
ggplot( data = acacia, mapping = aes(x = CIRC, y = HEIGHT)) +
  geom_point(size = 1, color = "purple", alpha = 0.4) +
  scale_x_log10() +
  scale_y_log10()
```

```
## Warning: Removed 4 rows containing missing values (geom_point).
```



Grouping data:

We have the information on experimental treatment in the column named "TREATMENT":

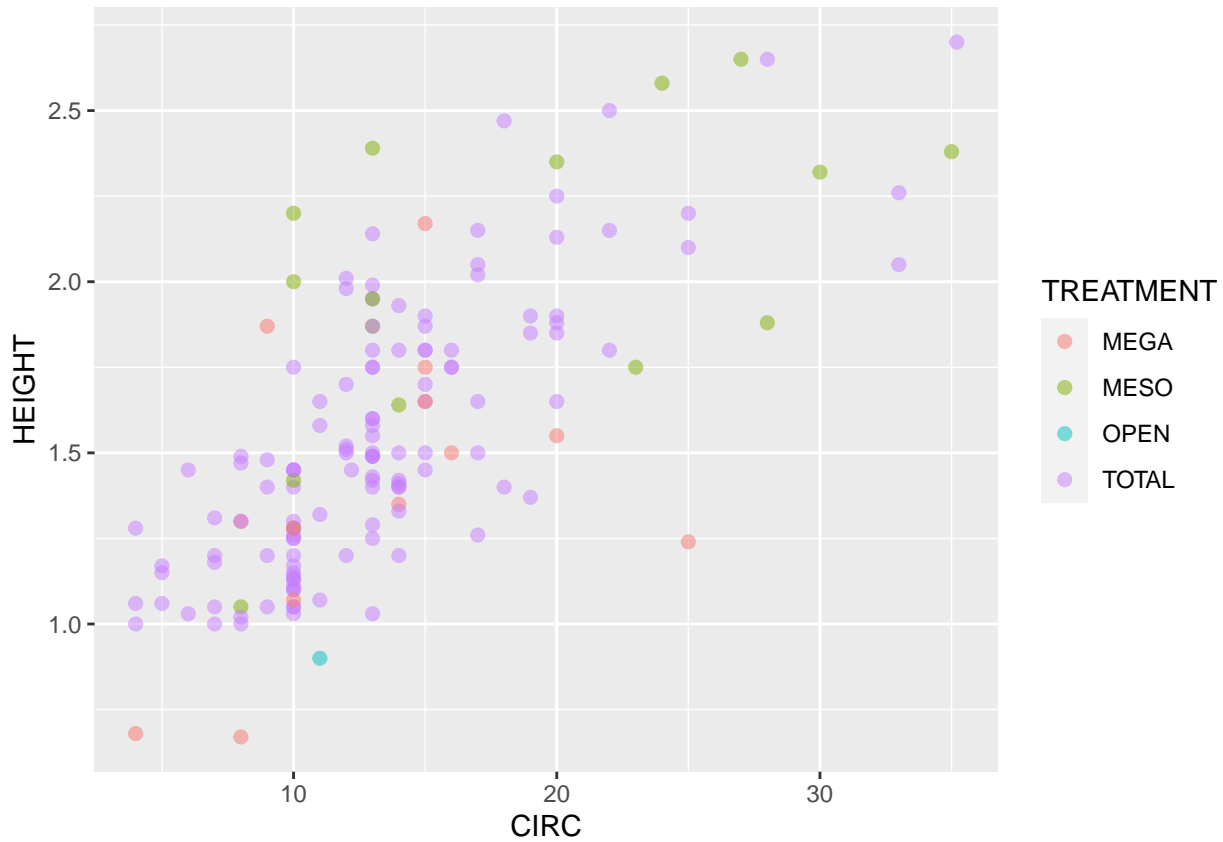
```
acacia$TREATMENT
```

```
## [1] "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "MESO"
## [10] "MESO" "MESO" "MESO" "MESO" "MESO" "MESO" "MESO" "MESO" "MESO" "MESO"
## [19] "MESO" "MESO" "OPEN" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL"
## [28] "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL"
## [37] "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL"
## [46] "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL"
## [55] "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL"
## [64] "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL"
## [73] "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL"
## [82] "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL"
## [91] "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL"
## [100] "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "MEGA"
## [109] "MEGA" "MEGA" "MEGA" "MEGA" "MEGA" "MEGA" "MEGA" "MEGA" "MEGA" "MEGA"
## [118] "MEGA" "MEGA" "MEGA" "MESO" "MESO" "MESO" "OPEN" "OPEN" "TOTAL"
## [127] "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL"
## [136] "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL"
## [145] "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL" "TOTAL"
## [154] "TOTAL" "TOTAL" "MESO" "MESO"
```

Let's add information about treatment to our plot:


```
ggplot(data = acacia, mapping = aes (x = CIRC, y = HEIGHT, color = TREATMENT)) +
  geom_point(size = 2, alpha = 0.5)
```

Warning: Removed 4 rows containing missing values (geom_point).

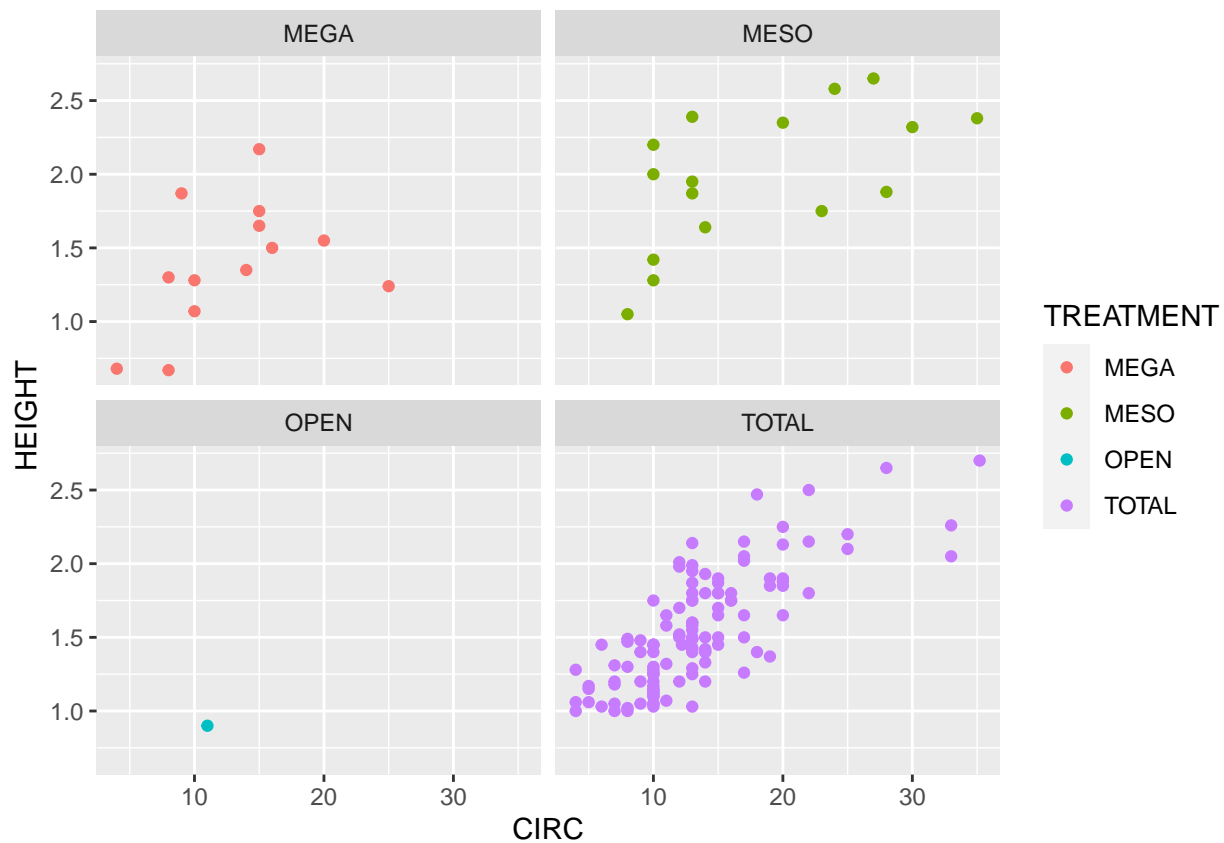


Subplots or facets

The function to create subplots by a third variable is called `facet_wrap()`

```
ggplot(data = acacia, mapping = aes(x = CIRC, y = HEIGHT, color = TREATMENT)) +
  geom_point() +
  facet_wrap(~TREATMENT)
```

Warning: Removed 4 rows containing missing values (geom_point).



A little interpretation of what is going on here

How to test these hypotheses?

Model fitting functions

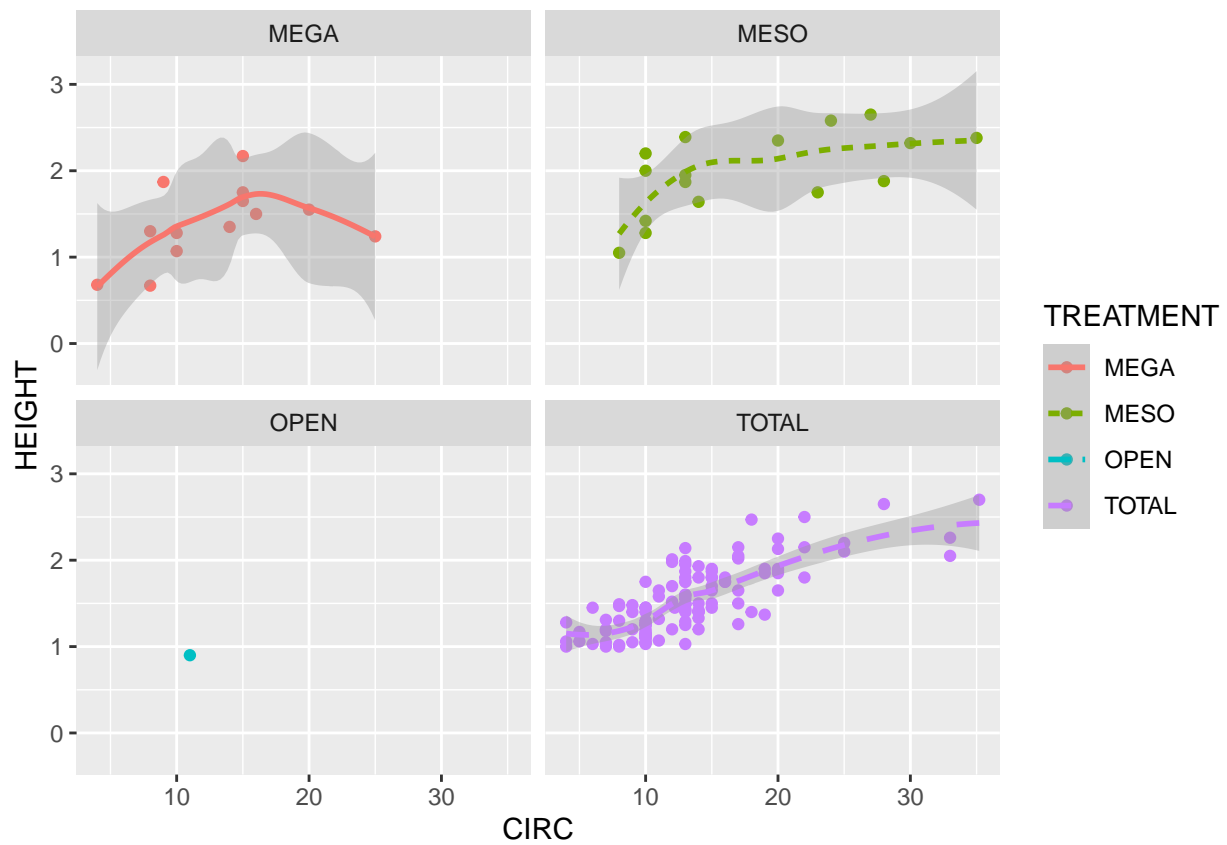
The `geom_smooth()` function allows to fit linear models to a set of points

```
ggplot(data = acacia, mapping = aes(x = CIRC, y = HEIGHT, color = TREATMENT, linetype = TREATMENT)) +
  geom_point() +
  geom_smooth(method = "loess") +
  facet_wrap(~TREATMENT)
```

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

```
## Warning: Removed 4 rows containing non-finite values (stat_smooth).
```

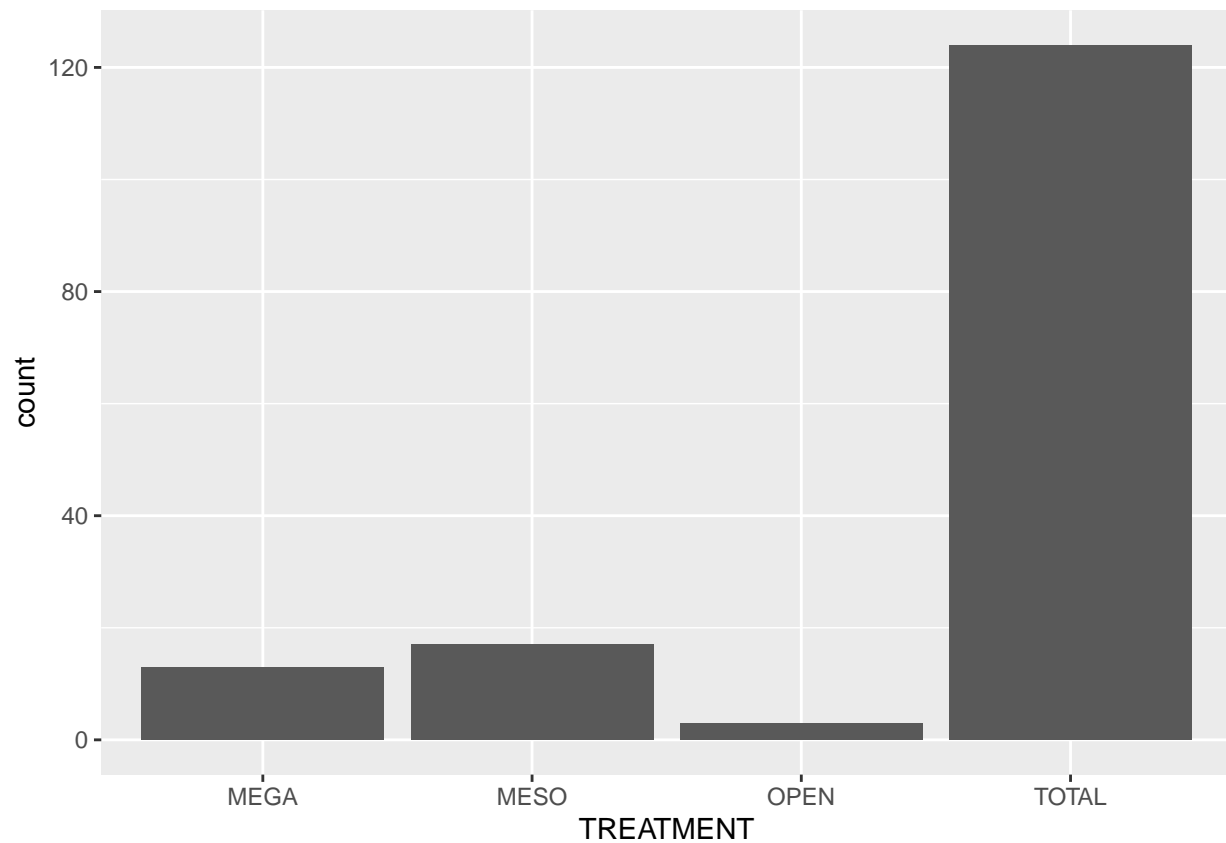
```
## Warning: Removed 4 rows containing missing values (geom_point).
```



Histograms and barplots

For bar plots we use the `geom_bar()` function:

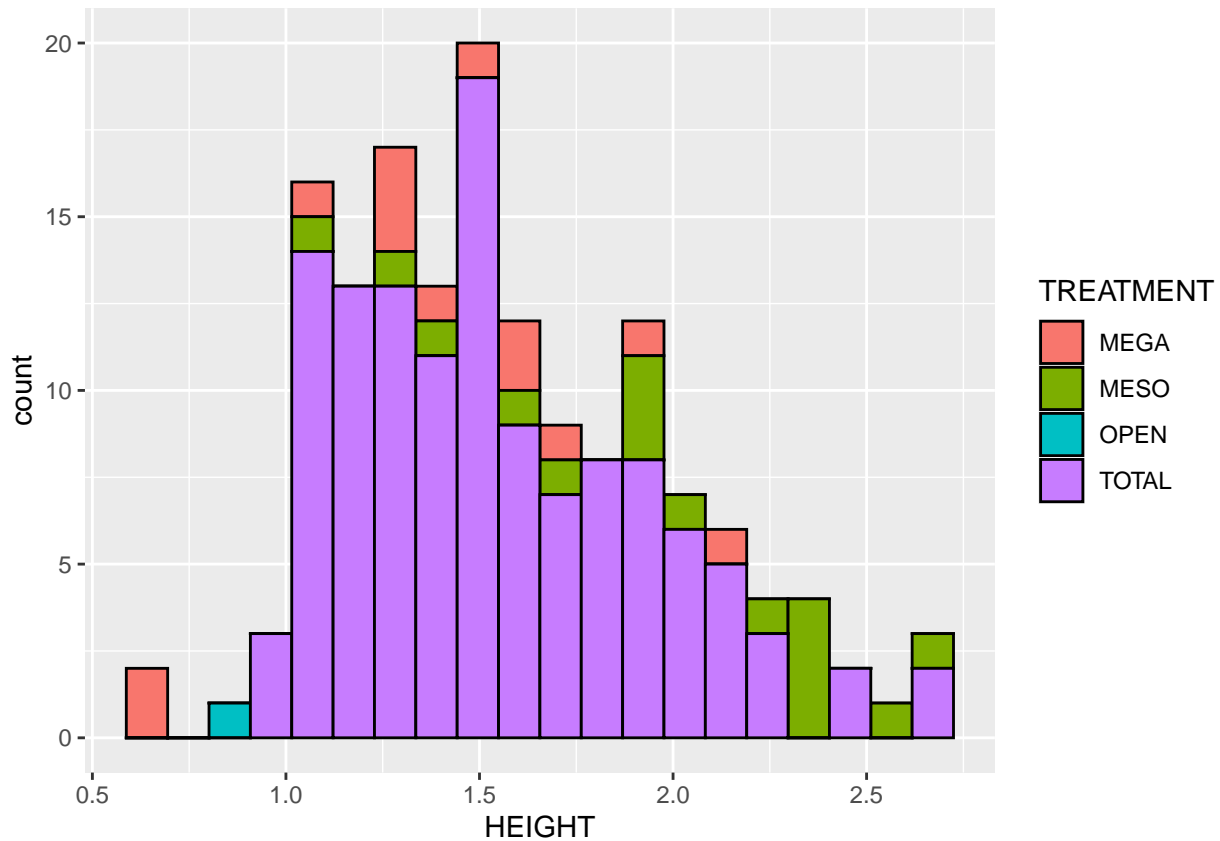
```
ggplot(data = acacia, aes(x = TREATMENT)) +  
  geom_bar()
```



If we want to see the distribution of a continuous variable we use the `geom_histogram()` function:

```
ggplot(data = acacia, mapping = aes(x = HEIGHT, fill = TREATMENT)) +  
  geom_histogram(bins = 20, color = "black")
```

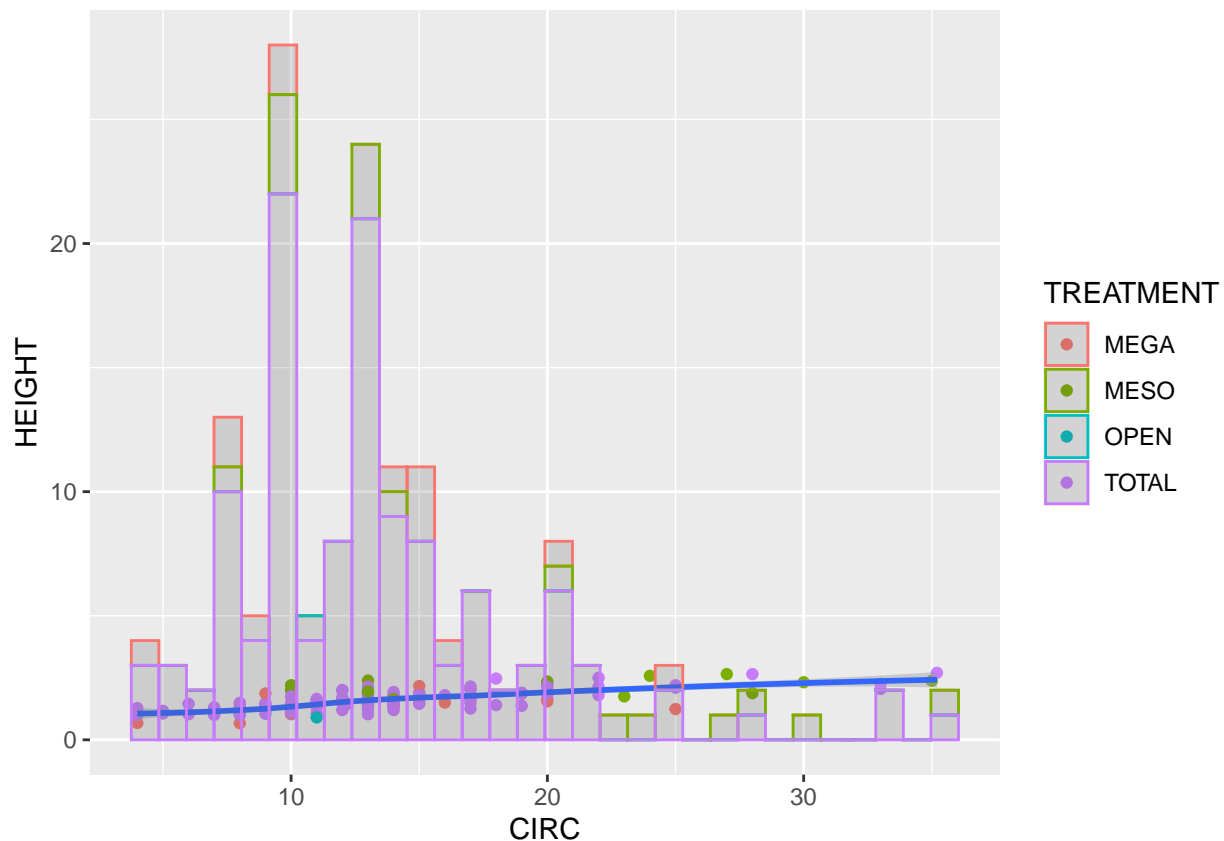
```
## Warning: Removed 4 rows containing non-finite values (stat_bin).
```



Layer multiple data from the same or different data sets

```
ggplot() +
  geom_point(data = acacia,
             mapping = aes(x = CIRC, y = HEIGHT,
                           color = TREATMENT)) +
  geom_smooth(data = acacia,
             mapping = aes(x = CIRC, y = HEIGHT)) +
  geom_histogram(data = acacia,
                mapping = aes(x = CIRC, color = TREATMENT), alpha = 0.2)
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
## Warning: Removed 4 rows containing non-finite values (stat_smooth).
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Warning: Removed 4 rows containing non-finite values (stat_bin).
## Warning: Removed 4 rows containing missing values (geom_point).
```



Save images as files

```
ggsave(filename = "acacia_by_treatment.pdf", width = 5, height = 3)

## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
## Warning: Removed 4 rows containing non-finite values (stat_smooth).
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Warning: Removed 4 rows containing non-finite values (stat_bin).
## Warning: Removed 4 rows containing missing values (geom_point).
```

Exercise 1

```
ggplot(data = acacia, mapping = aes(x = CIRC, y = AXIS1, color = ANT)) +
  geom_point() +
  labs(x = "Circumference", y = "Canopy Diameter") +
  facet_wrap(~ANT) +
  geom_smooth()

## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
## Warning: Removed 4 rows containing non-finite values (stat_smooth).
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : span too small. fewer data values than degrees of freedom.
```

```

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : at 9.975

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : radius 0.000625

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : all data on boundary of neighborhood. make span bigger

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : pseudoinverse used at 9.975

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : neighborhood radius 0.025

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : reciprocal condition number 1

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : at 15.025

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : radius 0.000625

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : all data on boundary of neighborhood. make span bigger

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : There are other near singularities as well. 0.000625

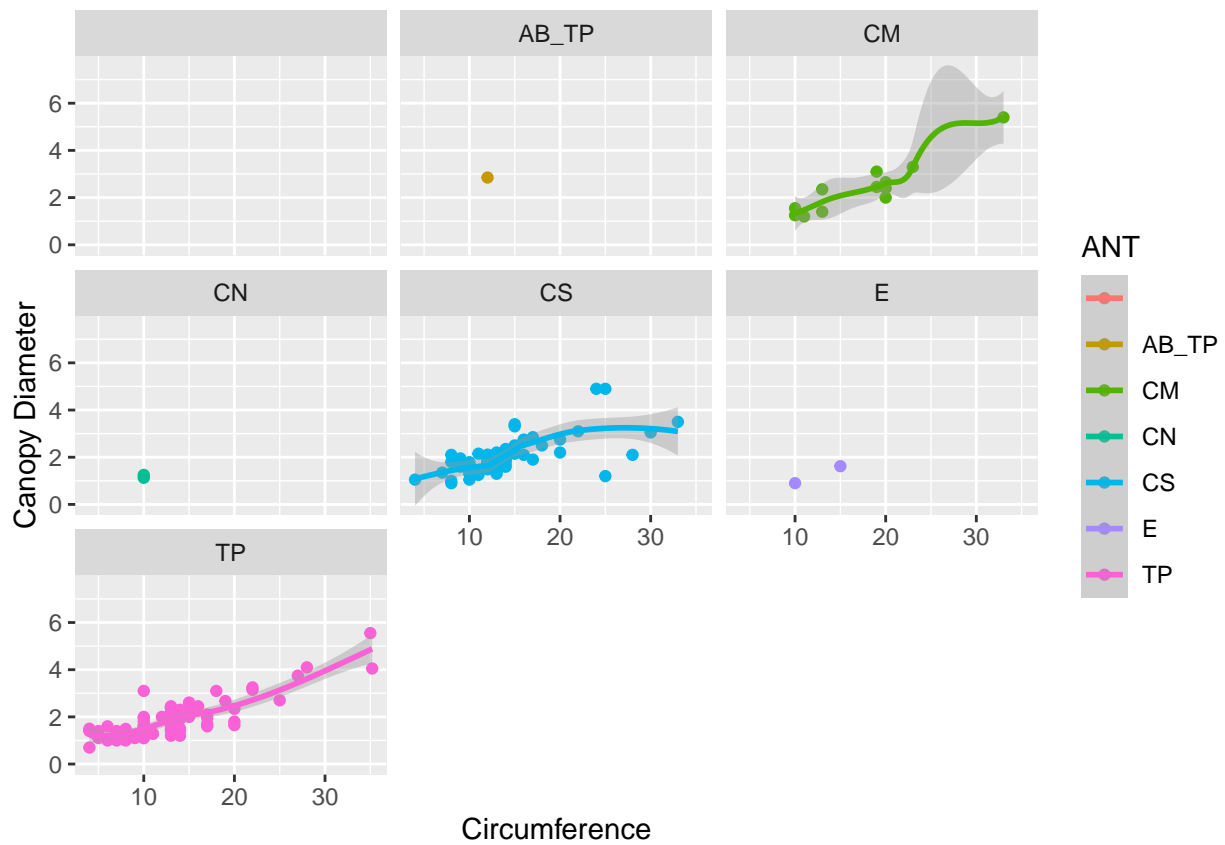
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : zero-width neighborhood. make span bigger

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : zero-width neighborhood. make span bigger

## Warning: Computation failed in `stat_smooth()`:
## NA/NaN/Inf in foreign function call (arg 5)

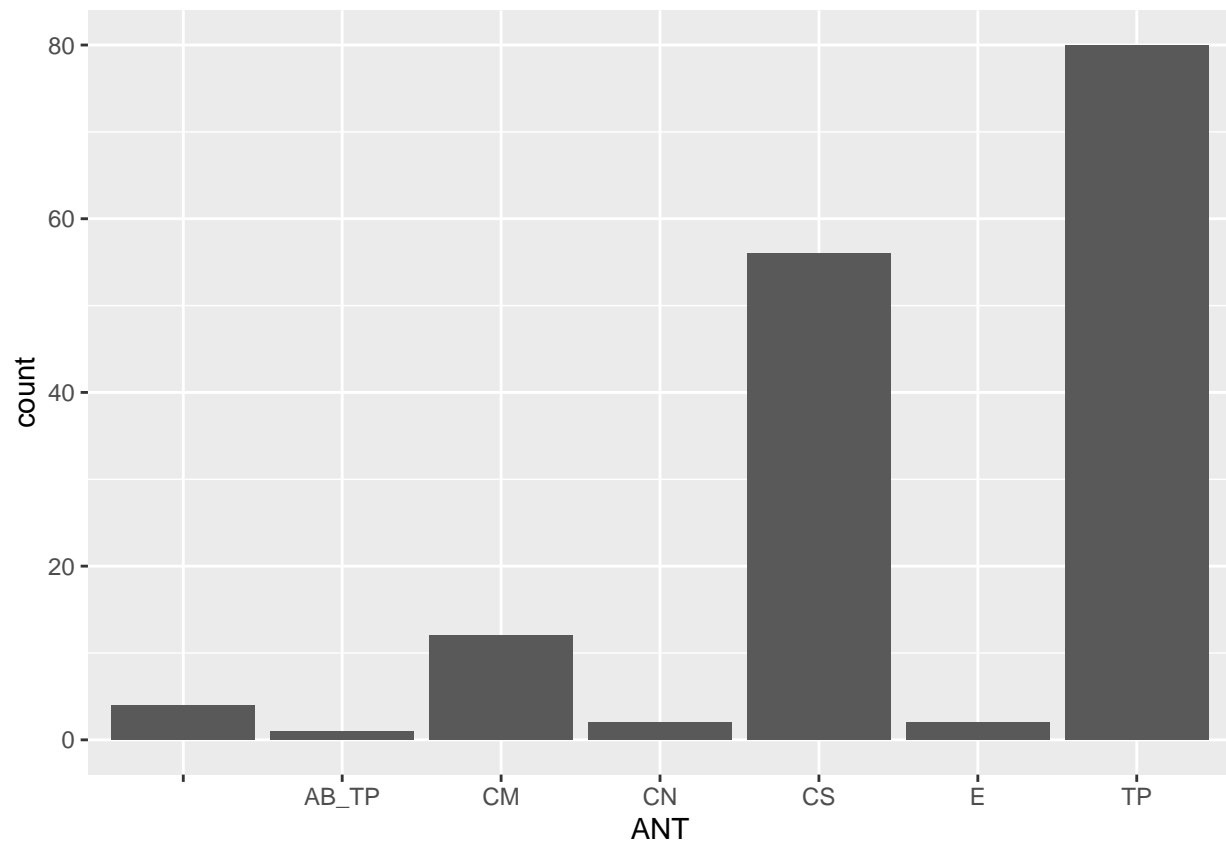
## Warning: Removed 4 rows containing missing values (geom_point).

```



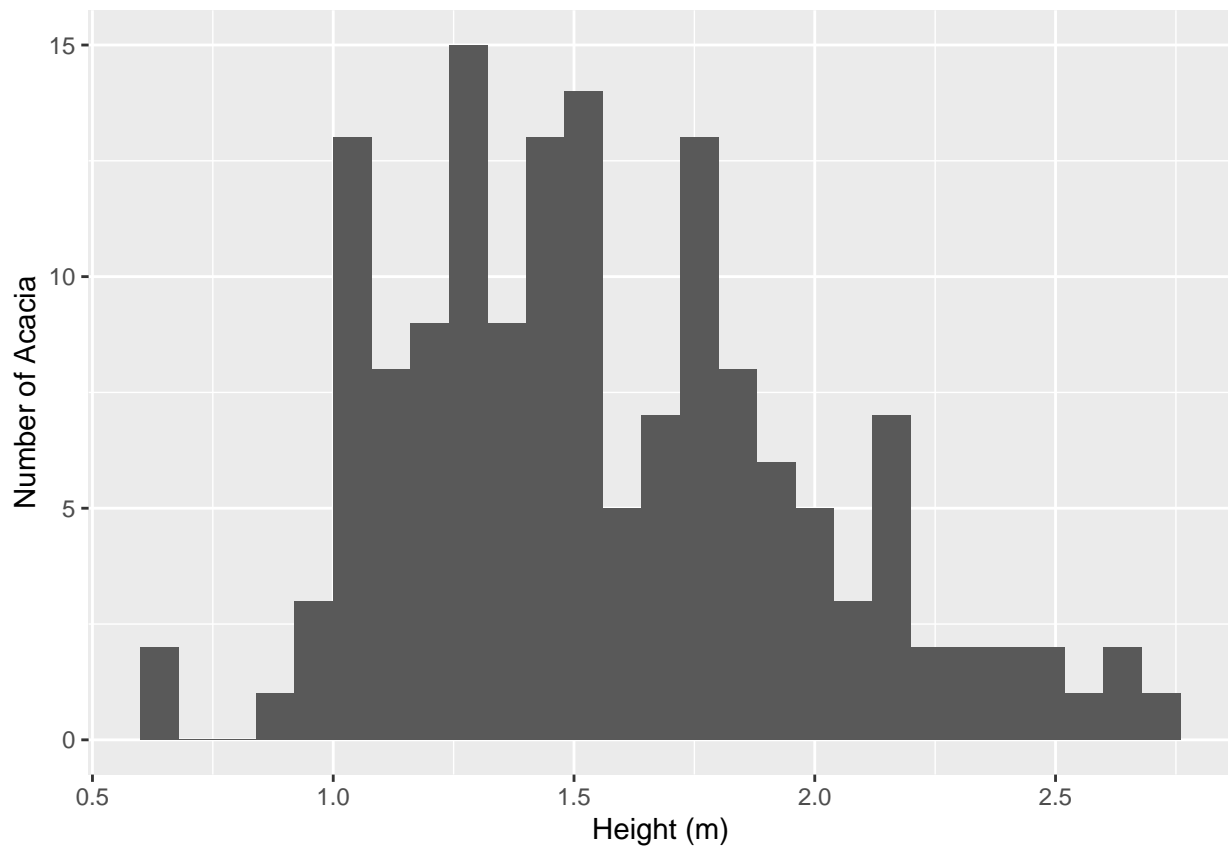
Exercise 2.

```
ggplot(data = acacia, mapping = aes(x = ANT)) +  
  geom_bar()
```

```
ggplot(data = acacia, mapping = aes(x = HEIGHT)) +  
  geom_histogram(binwidth = 0.08) +  
  labs(x = "Height (m)", y = "Number of Acacia")
```

```
## Warning: Removed 4 rows containing non-finite values (stat_bin).
```



```
ggplot(data = acacia) +  
  geom_histogram(mapping = aes(x = AXIS1),  
                 fill = "red", alpha = 0.3, bins = 10) +  
  geom_histogram(mapping = aes(x = AXIS2),  
                 fill = "black", alpha = 0.3, bins = 10) +  
  facet_wrap(~TREATMENT)
```

```
## Warning: Removed 4 rows containing non-finite values (stat_bin).  
## Removed 4 rows containing non-finite values (stat_bin).
```

