Monthly Report (00)

2015.09 Data Set

Sunday 25th October, 2015

Prepared for

Statistics for Physical and Engineering Sciences

by

Jamie Riggs, Ph.D.

Principal Statistician
Statistics for Physical and Engineering Sciences Institute

1   Introduction

The process of reporting monthly Sunspot numbers consists of submitting individual observer's daily counts for a specific month to the AAVSO Solar Section. These data are maintained in a SQL database. The monthly data then are extracted for analysis using the R statistics package (http://www.R-project.org/). This report is the portion of the analysis concerned with both the raw daily average counts and the data Accuracy, Consistency, and Completeness measures for a particular month. The checks are used to scrub or filter the data to assure only error-free data are used to determine the monthly sunspot number.

This report consists of four sections: the raw daily average counts (Section 2), the known data errors (Section 3), the processed counts using a Generalized Linear Mixed Model to produce the relative sunspot numbers (Section 4), and supporting information on the model construction (Section 5).

The raw daily average of counts consist of submitted counts from all observers who provided data in the particular month. These averaged counts are reported by the day of the month, and are either from data not scrubbed or corrected data. The table captions indicate which. The errors, if any, are reported according to type.

The Error Tables section contains reported errors on missing data, inconsistencies in year and month, inconsistencies in the reported day number (1-31), seeing coding errors, number of annual observations by observer, and inconsistencies between the reported Wolf number and the calculated Wolf number from the group counts and sunspot counts, among other errors that are given in that section.

The relative sunspot numbers R_a section contains the sunspot numbers after the submitted data are scrubbed and modeled by a Generalized Linear Mixed Model (GLMM). The GLMM is a statistical model that accounts for variation due to random effects and fixed effects. For the R_a model random effects include the AAVSO observer as these observers are a selection from all possible observers, and the fixed effects include seeing conditions at one of four possible levels. More details on GLMM are available in a paper on the sunspot counts research page. The paper title is *A Generalized Linear Mixed Model for Enumerated Sunspots*.

The supporting information for the model is provided for clarification.

2   Raw Daily Average Counts

The reported raw daily average counts have been checked for errors and inconsistencies, and no known errors are present. All observers whose submissions qualify through this month's scrubbing process are represented in Figure 1 and Table 1.
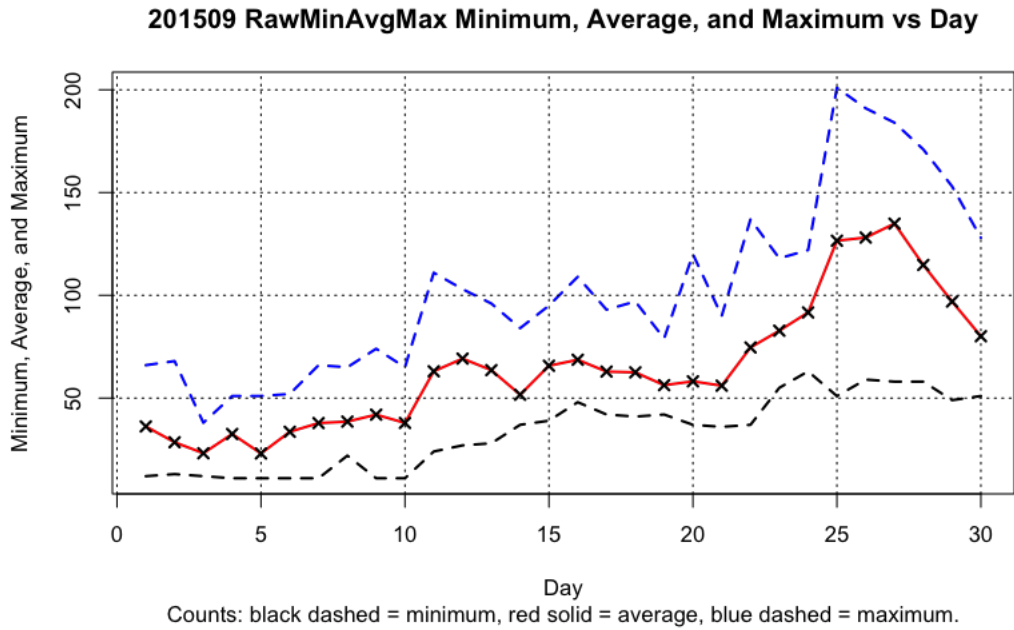
201509 RawMinAvgMax Minimum, Average, and Maximum vs Day

Counts: black dashed = minimum, red solid = average, blue dashed = maximum.

Figure 1: Raw average sunspot count by day of the month.

Table 1: 201509 Daily Raw Counts

Day	Submissions	Minimum	Average	Maximum
1.0000	38.0000	12.0000	36.2059	66.0000
2.0000	43.0000	13.0000	28.5128	68.0000
3.0000	36.0000	12.0000	23.1667	38.0000
4.0000	31.0000	11.0000	32.5333	51.0000
5.0000	37.0000	11.0000	23.0606	51.0000
6.0000	37.0000	11.0000	33.5882	52.0000
7.0000	41.0000	11.0000	37.8421	66.0000
8.0000	33.0000	22.0000	38.5625	65.0000
9.0000	34.0000	11.0000	41.9375	74.0000
10.0000	32.0000	11.0000	37.9000	65.0000
11.0000	37.0000	24.0000	63.0303	111.0000
12.0000	22.0000	27.0000	69.2000	103.0000
13.0000	33.0000	28.0000	63.6000	96.0000
14.0000	38.0000	37.0000	51.6316	84.0000
15.0000	40.0000	39.0000	65.7297	95.0000
16.0000	36.0000	48.0000	68.6364	109.0000
17.0000	38.0000	42.0000	62.8857	93.0000
18.0000	40.0000	41.0000	62.5135	97.0000
19.0000	35.0000	42.0000	56.2857	79.0000
20.0000	41.0000	37.0000	58.1471	120.0000
21.0000	42.0000	36.0000	56.0263	90.0000
22.0000	40.0000	37.0000	74.6176	137.0000
23.0000	33.0000	55.0000	82.7812	118.0000
24.0000	38.0000	63.0000	91.6667	122.0000
25.0000	35.0000	51.0000	126.5625	201.0000
26.0000	33.0000	59.0000	128.1034	191.0000
27.0000	37.0000	58.0000	134.8710	184.0000
28.0000	34.0000	58.0000	114.7667	171.0000
29.0000	32.0000	49.0000	96.9643	153.0000
30.0000	34.0000	51.0000	80.1000	128.0000

3 Error Tables

Data are for the month of September 2015. No errors were found, and hence no errors are reported.

4 Relative Sunspot Numbers

All data errors, if any, have been corrected prior to determining the following relative sunspot numbers. A Generalized Linear Mixed Model (GLMM) was constructed to provide monthly sunspot numbers (see Table 2). The GLMM treats observer as a random effect, with year, month, seeing conditions, observer rank, and dual submission to both AAVSO and SILSO as fixed effects.

Figure 2 shows the monthly R_a numbers for the years and months (ym) in Table 2. The solid cyan curve that connects the cyan X's are the GLMM model estimates given in 2. The dotted black curves on either side of the cyan curve depict a 99% confidence band about the GLMM estimates. The confidence band uses the large sample approximation based on the Gaussian distribution. The dashed red curve connecting the red O's are the SILSO values for the monthly sequence.

The tan box plots for each month are the actual observations submitted by the AAVSO observers. The heavy solid lines approximately midway in the boxes represent the count medians. The box of the box plot represents the InterQuartile Range (IQR), which depicts from the 25th through the 75th quartiles. The lower and upper whiskers extend 1.5 times the IQR below the 25th quartile, and 1.5 times the IQR above the 75th quartile. The black circles below and above the whiskers traditionally are considered outliers, but with GLMM modeling, they are observations that comprise overdispersion. Overdispersion skews the counts data from a true Poisson distribution. The GLMM adjusts for this overdispersion.

Table 2: Year Month (ym) Relative Sunspot Numbers with 99% CI

ym	Ra	lci99	uci99
2010.05	22.8239	22.3070	23.3408
2010.06	18.3713	17.8851	18.8575
2010.07	20.1545	19.7131	20.5959
2010.08	19.2489	18.7845	19.7132
2010.09	23.2050	22.7138	23.6963
2010.10	23.2186	22.7219	23.7153
2010.11	25.3377	24.7696	25.9059
2010.12	23.8693	23.1911	24.5476
2011.01	72.7108	71.0894	74.3321
2011.02	61.4777	60.0951	62.8602
2011.03	69.9583	68.5247	71.3919
2011.04	76.3686	74.7385	77.9987
2011.05	78.1698	76.6069	79.7327
2011.06	66.3347	64.9444	67.7250

Table 2: Year Month (ym) Relative Sunspot Numbers with 99% CI

ym	Ra	lci99	uci99
2011.07	70.9354	69.3826	72.4882
2011.08	70.9372	69.5536	72.3208
2011.09	82.8935	81.8112	83.9758
2011.10	82.5196	81.1152	83.9241
2011.11	88.5558	86.6889	90.4228
2011.12	80.9110	79.1599	82.6621
2012.01	74.7928	73.3241	76.2614
2012.02	61.1625	59.8769	62.4480
2012.03	72.4824	71.1949	73.7698
2012.04	76.3371	74.1641	78.5100
2012.05	82.1732	80.7432	83.6032
2012.06	69.5063	68.2758	70.7368
2012.07	75.3060	74.0441	76.5679
2012.08	71.7797	70.5861	72.9733
2012.09	83.7677	82.3433	85.1921
2012.10	84.9399	83.3428	86.5371
2012.11	92.3115	90.4731	94.1499
2012.12	81.9973	80.2694	83.7253
2013.01	84.5264	82.9447	86.1081
2013.02	70.5585	69.1857	71.9312
2013.03	79.5507	78.0573	81.0441
2013.04	88.8701	87.3769	90.3633
2013.05	89.9184	88.3729	91.4639
2013.06	76.6470	75.2870	78.0069
2013.07	80.9156	79.6547	82.1765
2013.08	79.1775	77.9432	80.4118
2013.09	91.6247	90.0491	93.2003
2013.10	91.4540	89.8408	93.0672
2013.11	99.4295	97.3729	101.4861
2013.12	90.6467	88.8283	92.4651
2014.01	100.4383	98.3344	102.5422
2014.02	84.1193	82.5301	85.7086
2014.03	97.9539	96.3077	99.6002
2014.04	108.0390	106.2114	109.8665
2014.05	108.9511	107.2179	110.6844
2014.06	92.9267	91.4215	94.4319
2014.07	99.5124	97.8979	101.1269
2014.08	96.7716	95.3188	98.2243

Table 2: Year Month (ym) Relative Sunspot Numbers with 99% CI

| ym	Ra	lci99	uci99
2014.09	113.0436	111.2464	114.8408
2014.10	112.5020	110.6279	114.3762
2014.11	122.9533	120.6270	125.2796
2014.12	109.4638	107.1226	111.8051
2015.01	66.6693	65.3879	67.9508
2015.02	55.7556	54.4733	57.0379
2015.03	63.8515	62.6910	65.0120
2015.04	71.2618	70.0077	72.5159
2015.05	71.5548	70.4092	72.7004
2015.06	61.4259	60.3854	62.4664
2015.07	64.3805	63.2884	65.4727
2015.08	64.1540	63.1076	65.2005
2015.09	74.7308	73.5235	75.9381
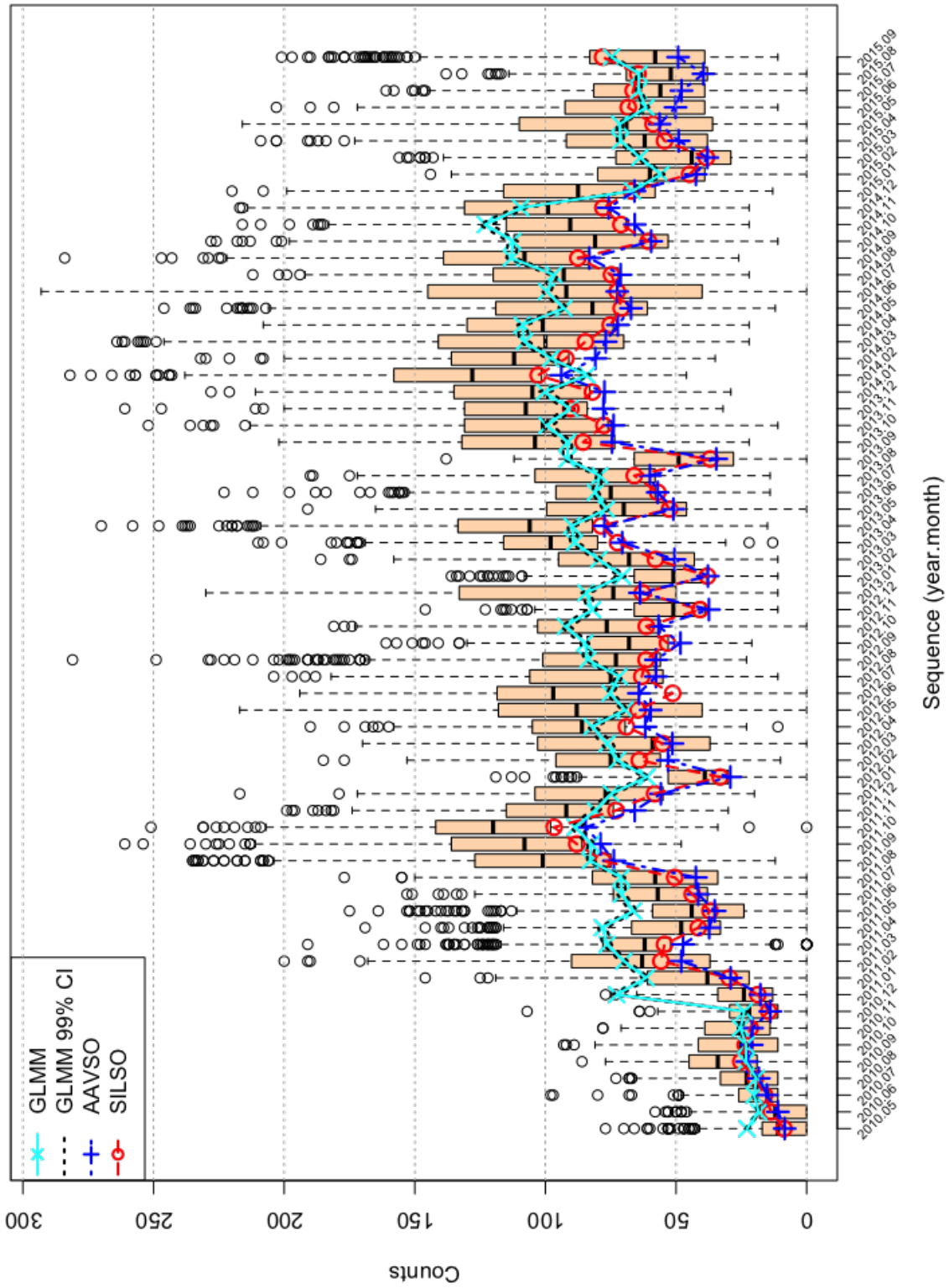
Figure 2: GLMM fitted data for R_a.

The GLMM parameter estimates and measures of importance in the determining the monthly R_a values are given in Table 3. The parameter estimates and levels of statistical significance are determined for the residual error size combined with the observer random effect error size. Thus, the parameter estimates are adjusted for the random effect of observer. The significance level is set at 0.05. Any Pr($>$|z|) values equal to or less than 0.05 are considered statistically significant.

Table 3: 201509 Parameter Estimates

	Estimate	Std. Error	t-value	Pr($>$|t|)
(Intercept)	3.1551	0.0445	70.9109	0.0000
seeF	-0.1829	0.0076	-24.1790	0.0000
seeG	-0.0946	0.0066	-14.3761	0.0000
seeP	-0.2935	0.0112	-26.3066	0.0000
r1000B	-0.0559	0.0829	-0.6745	0.5000
r1500C	0.0363	0.1266	0.2866	0.7744
r2000D	0.0848	0.1544	0.5490	0.5830
r2500E	0.0001	0.1049	0.0012	0.9991
r3000F	0.0612	0.1021	0.5993	0.5490
r3500G	0.1198	0.1528	0.7842	0.4329
r5000H	-0.1108	0.2113	-0.5245	0.5999
silsoy	0.1208	0.0736	1.6407	0.1009
year2011	1.2236	0.0153	79.8253	0.0000
year2012	1.2409	0.0153	81.2088	0.0000
year2013	1.3381	0.0152	87.8089	0.0000
year2014	1.5279	0.0151	100.9137	0.0000
year2015	1.1158	0.0160	69.9054	0.0000
mon2	-0.1880	0.0128	-14.6898	0.0000
mon3	-0.0532	0.0116	-4.5819	0.0000
mon4	0.0482	0.0117	4.1287	0.0000
mon5	0.0590	0.0110	5.3417	0.0000
mon6	-0.1096	0.0116	-9.4674	0.0000
mon7	-0.0464	0.0112	-4.1522	0.0000
mon8	-0.0630	0.0111	-5.6993	0.0000
mon9	0.0933	0.0106	8.7975	0.0000
mon10	0.0951	0.0115	8.2527	0.0000
mon11	0.1882	0.0120	15.6708	0.0000
mon12	0.0827	0.0125	6.6205	0.0000

The year effect levels are given as year2011, year2012, and year2013. The yearly effect is significant as $Pr(> |z|) < 0.05$. So the year in which the observations are made is commensurate with the expected rise toward and anticipated sunspot number maximum. Similarly, the monthly effect, denoted as mon2 through mon12, is significant at the 0.05 level.

The seeing conditions account for a significant amount of deviation in sunspot numbers. The

seeing conditions are denoted as seeF (Fair), seeG (Good), and seeP (Poor), and are significant at the 0.05 level. Therefore, seeing conditions influence the reported sunspot numbers, as intuition anticipates.

The level of observer experience (denoted r1000B through r5000H, which is least to most experience) is not significant at the 0.05 significance level. It therefore does not contribute to changes in the monthly sunspot numbers.

Whether an observer contributes counts to the SILSO as well as the AAVSO (silsoy) is not significant at the 0.05 level, and hence we conclude that those observers who contribution to both institutions tend to differ from those observers contributing only to the AAVSO.

5   Supporting Information

Table 4: 201509 Summary of Sunspot Numbers

| obs	jd	year	mon	day
ARAG : 1915	Min. :1721096	Min. :2010	Min. : 1.00	Min. : 1.00
CHAG : 1733	1st Qu.:2455873	1st Qu.:2011	1st Qu.: 4.00	1st Qu.: 8.00
BRAB : 1703	Median :2456363	Median :2013	Median : 7.00	Median :16.00
BROB : 1533	Mean :2455959	Mean :2013	Mean : 6.61	Mean :15.73
DUBF : 1448	3rd Qu.:2456818	3rd Qu.:2014	3rd Qu.: 9.00	3rd Qu.:23.00
HOWR : 1408	Max. :2457296	Max. :2015	Max. :12.00	Max. :31.00
(Other):35527				

Table 5: Summary of Sunspot Numbers

see	g	s	w	r	silso
E: 8337	Min. : 0.000	Min. : 0.00	Min. : 0.00	0000A :19587	n:30720
F:13768	1st Qu.: 3.000	1st Qu.: 12.00	1st Qu.: 44.00	3000F : 7220	y:14547
G:19504	Median : 5.000	Median : 24.00	Median : 73.00	2500E : 5838	
P: 3658	Mean : 4.735	Mean : 29.96	Mean : 77.31	3500G : 3436	
	3rd Qu.: 6.000	3rd Qu.: 43.00	3rd Qu.:106.00	1000B : 3294	
	Max. :18.000	Max. :204.00	Max. :293.00	1500C : 2695	
				(Other): 3197	
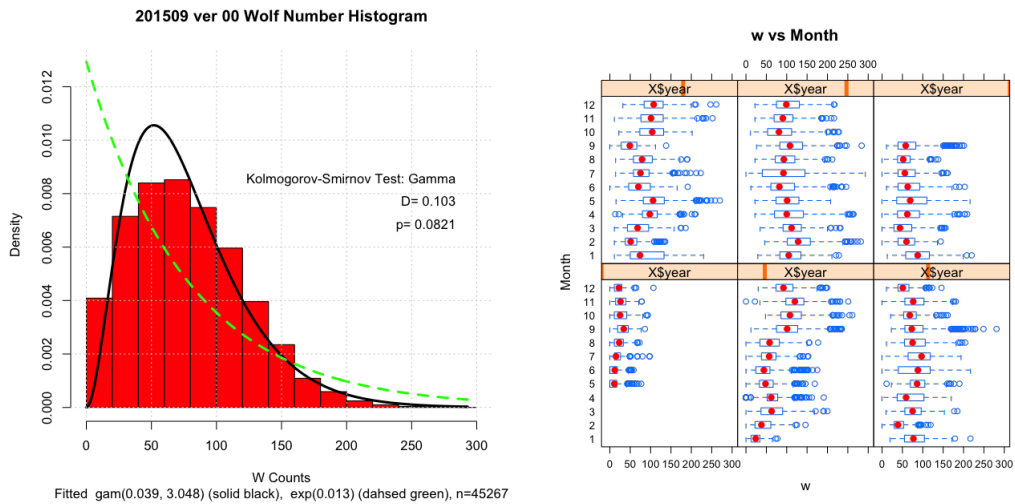
(a) Observed sunspot count histogram.

(b) Box plot of sunspot count by year and month.
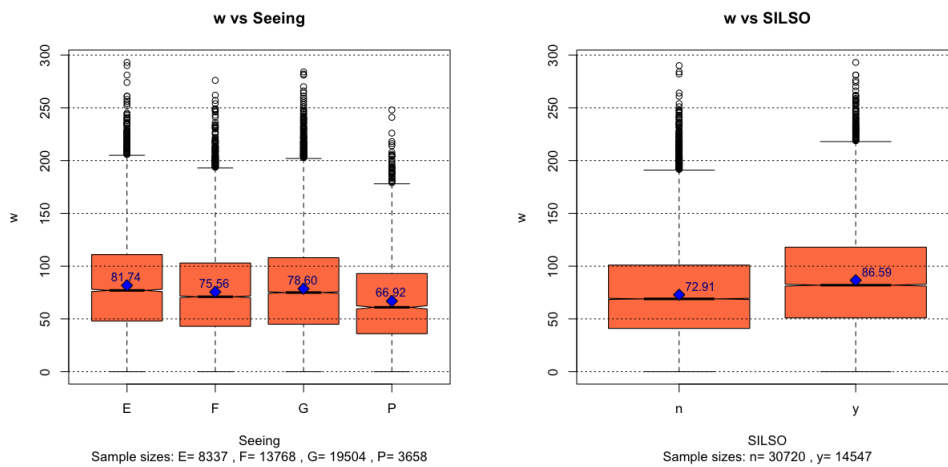
(c) Box plot of sunspot count by seeing condition.
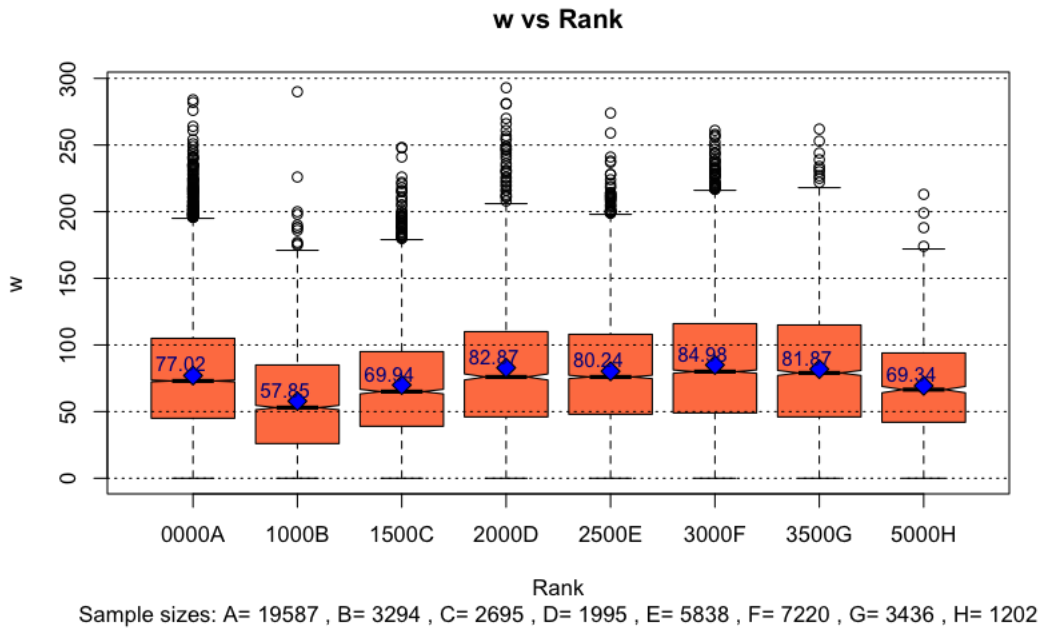

(d) Box plot of sunspot count submitted to AAVSO and SILSO.

Figure 3:

Figure 4: Box plot of sunspot count by rank.