Monthly Report (00)

2015.08 Data Set

Friday 18th September, 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.
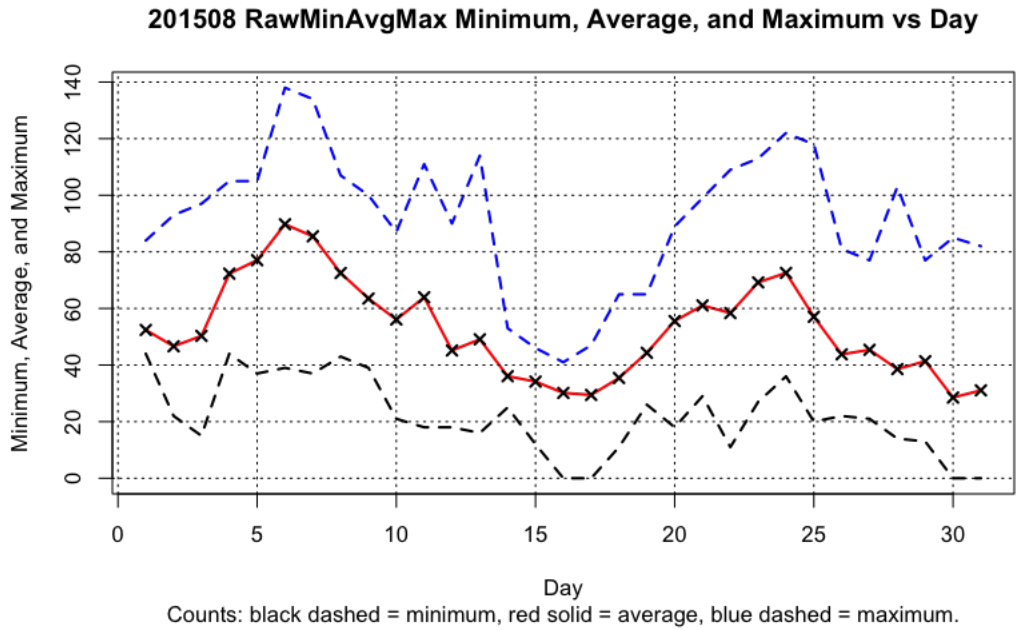
201508 RawMinAvgMax Minimum, Average, and Maximum vs Day

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

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

Table 1: 201508 Daily Raw Counts

Day	Submissions	Minimum	Average	Maximum
1.0000	36.0000	44.0000	52.4118	84.0000
2.0000	42.0000	22.0000	46.6316	93.0000
3.0000	37.0000	15.0000	50.2973	97.0000
4.0000	35.0000	44.0000	72.2941	105.0000
5.0000	41.0000	37.0000	77.0513	105.0000
6.0000	32.0000	39.0000	89.7500	138.0000
7.0000	38.0000	37.0000	85.4857	134.0000
8.0000	39.0000	43.0000	72.5294	107.0000
9.0000	37.0000	39.0000	63.5000	100.0000
10.0000	36.0000	21.0000	56.0882	87.0000
11.0000	36.0000	18.0000	63.9429	111.0000
12.0000	36.0000	18.0000	45.1515	90.0000
13.0000	34.0000	16.0000	49.0938	114.0000
14.0000	34.0000	25.0000	36.0606	53.0000
15.0000	35.0000	12.0000	34.1515	46.0000
16.0000	35.0000	0.0000	30.1613	41.0000
17.0000	32.0000	0.0000	29.4333	47.0000
18.0000	32.0000	11.0000	35.4333	65.0000
19.0000	38.0000	26.0000	44.3333	65.0000
20.0000	38.0000	18.0000	55.5405	89.0000
21.0000	35.0000	29.0000	61.0588	99.0000
22.0000	43.0000	11.0000	58.3684	109.0000
23.0000	37.0000	27.0000	69.2059	113.0000
24.0000	36.0000	36.0000	72.5625	122.0000
25.0000	38.0000	20.0000	57.0556	118.0000
26.0000	42.0000	22.0000	43.7949	81.0000
27.0000	35.0000	21.0000	45.3871	77.0000
28.0000	40.0000	14.0000	38.5676	103.0000
29.0000	38.0000	13.0000	41.3611	77.0000
30.0000	37.0000	0.0000	28.5152	85.0000
31.0000	35.0000	0.0000	31.0606	82.0000

3   Error Tables

Data are for the month of August 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	22.7256	22.2123	23.2388
2010.06	18.2893	17.8060	18.7726
2010.07	20.0663	19.6277	20.5049
2010.08	19.1698	18.7080	19.6315
2010.09	23.5887	23.0910	24.0864
2010.10	23.1859	22.6905	23.6813
2010.11	25.3122	24.7460	25.8783
2010.12	23.8472	23.1703	24.5241
2011.01	72.3026	70.6902	73.9151
2011.02	61.1827	59.8098	62.5555
2011.03	69.5793	68.1549	71.0038
2011.04	75.8873	74.2683	77.5063
2011.05	77.7440	76.1926	79.2955
2011.06	65.9845	64.6047	67.3644

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

ym	Ra	lci99	uci99
2011.07	70.5517	69.0080	72.0953
2011.08	70.5493	69.1728	71.9259
2011.09	84.1019	83.0034	85.2004
2011.10	82.3544	80.9542	83.7546
2011.11	88.3843	86.5228	90.2457
2011.12	80.7568	79.0093	82.5042
2012.01	74.5432	73.0808	76.0056
2012.02	61.0111	59.7301	62.2922
2012.03	72.2627	70.9800	73.5455
2012.04	76.1315	73.9792	78.2837
2012.05	81.8911	80.4669	83.3154
2012.06	69.2917	68.0663	70.5172
2012.07	75.0728	73.8158	76.3298
2012.08	71.5767	70.3886	72.7647
2012.09	85.2550	83.8076	86.7024
2012.10	84.9622	83.3647	86.5596
2012.11	92.3362	90.4967	94.1758
2012.12	82.0322	80.3027	83.7616
2013.01	84.2814	82.7048	85.8581
2013.02	70.3645	68.9943	71.7347
2013.03	79.3063	77.8177	80.7950
2013.04	88.5494	87.0637	90.0351
2013.05	89.6800	88.1418	91.2182
2013.06	76.4317	75.0782	77.7852
2013.07	80.7003	79.4455	81.9552
2013.08	78.9607	77.7309	80.1904
2013.09	93.2479	91.6439	94.8518
2013.10	91.5418	89.9280	93.1557
2013.11	99.5258	97.4681	101.5835
2013.12	90.7013	88.8795	92.5232
2014.01	100.1543	98.0556	102.2530
2014.02	83.9331	82.3474	85.5187
2014.03	97.6722	96.0293	99.3151
2014.04	107.6502	105.8334	109.4669
2014.05	108.6433	106.9161	110.3705
2014.06	92.6600	91.1602	94.1598
2014.07	99.2301	97.6204	100.8398
2014.08	96.5064	95.0600	97.9529

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

| ym	Ra	lci99	uci99
2014.09	115.0717	113.2460	116.8974
2014.10	112.5768	110.7015	114.4520
2014.11	123.0347	120.7040	125.3655
2014.12	109.5593	107.2122	111.9064
2015.01	67.8801	66.5748	69.1854
2015.02	56.7925	55.4870	58.0980
2015.03	65.0144	63.8344	66.1944
2015.04	72.4707	71.1966	73.7447
2015.05	72.8384	71.6745	74.0023
2015.06	62.5264	61.4667	63.5862
2015.07	65.5711	64.4611	66.6811
2015.08	65.3387	64.2748	66.4027
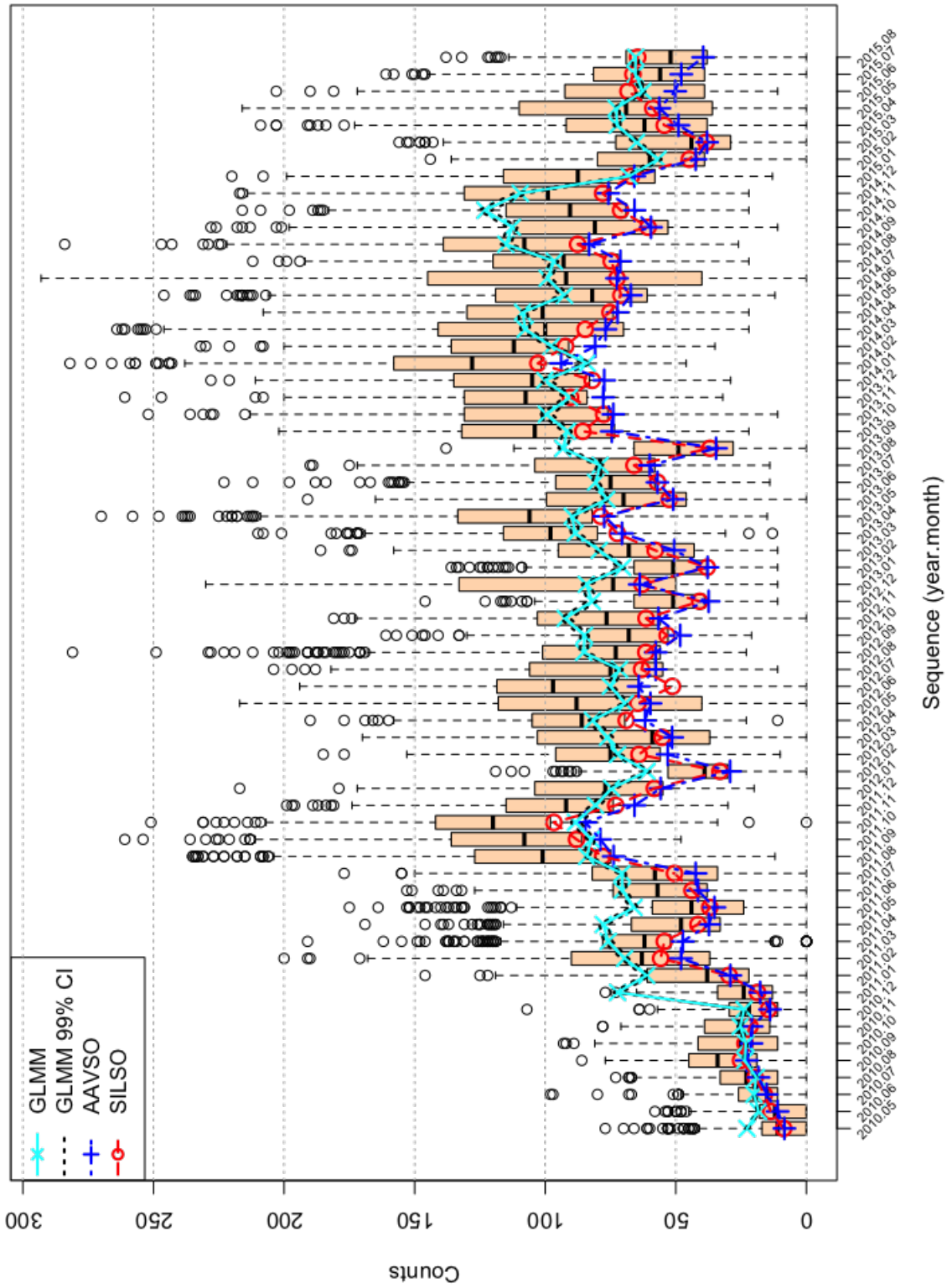
Figure 2: GLMM fitted data for R_a.

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: 201508 Parameter Estimates

	Estimate	Std. Error	t-value	Pr(>|t|)
(Intercept)	3.1492	0.0445	70.8062	0.0000
seeF	-0.1835	0.0076	-24.0757	0.0000
seeG	-0.0943	0.0066	-14.2229	0.0000
seeP	-0.2921	0.0112	-26.0439	0.0000
r1000B	-0.0541	0.0828	-0.6535	0.5135
r1500C	0.0382	0.1265	0.3020	0.7627
r2000D	0.0839	0.1543	0.5440	0.5865
r2500E	0.0005	0.1048	0.0050	0.9960
r3000F	0.0617	0.1020	0.6044	0.5456
r3500G	0.1212	0.1527	0.7938	0.4273
r5000H	-0.1089	0.2112	-0.5156	0.6061
silsoy	0.1214	0.0736	1.6500	0.0990
year2011	1.2231	0.0153	79.8348	0.0000
year2012	1.2428	0.0153	81.3662	0.0000
year2013	1.3403	0.0152	87.9858	0.0000
year2014	1.5303	0.0151	101.1050	0.0000
year2015	1.1391	0.0162	70.3273	0.0000
mon2	-0.1876	0.0128	-14.6599	0.0000
mon3	-0.0532	0.0116	-4.5848	0.0000
mon4	0.0472	0.0117	4.0486	0.0001
mon5	0.0588	0.0110	5.3271	0.0000
mon6	-0.1098	0.0116	-9.4894	0.0000
mon7	-0.0466	0.0112	-4.1702	0.0000
mon8	-0.0631	0.0111	-5.7083	0.0000
mon9	0.1138	0.0109	10.4385	0.0000
mon10	0.0988	0.0115	8.5628	0.0000
mon11	0.1917	0.0120	15.9595	0.0000
mon12	0.0864	0.0125	6.9197	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: 201508 Summary of Sunspot Numbers

obs	jd	year	mon	day
ARAG : 1885	Min. :1721096	Min. :2010	Min. : 1.000	Min. : 1.00
CHAG : 1704	1st Qu.:2455865	1st Qu.:2011	1st Qu.: 4.000	1st Qu.: 8.00
BRAB : 1677	Median :2456346	Median :2013	Median : 7.000	Median :16.00
BROB : 1505	Mean :2455937	Mean :2013	Mean : 6.569	Mean :15.74
DUBF : 1420	3rd Qu.:2456798	3rd Qu.:2014	3rd Qu.: 9.000	3rd Qu.:23.00
HOWR : 1379	Max. :2457266	Max. :2015	Max. :12.000	Max. :31.00
(Other):34948				

Table 5: Summary of Sunspot Numbers

| see	g	s	w	r	silso
E: 8182	Min. : 0.000	Min. : 0.00	Min. : 0.0	0000A :19283	n:30210
F:13538	1st Qu.: 3.000	1st Qu.: 12.00	1st Qu.: 44.0	3000F : 7091	y:14308
G:19187	Median : 5.000	Median : 24.00	Median : 73.0	2500E : 5742	
P: 3611	Mean : 4.749	Mean : 30.01	Mean : 77.5	3500G : 3381	
	3rd Qu.: 6.000	3rd Qu.: 43.00	3rd Qu.:106.0	1000B : 3229	
	Max. :18.000	Max. :204.00	Max. :293.0	1500C : 2635	
				(Other): 3157	
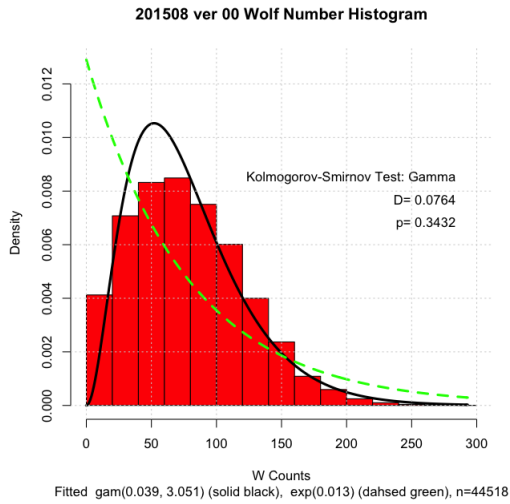
(a) Observed sunspot count histogram.
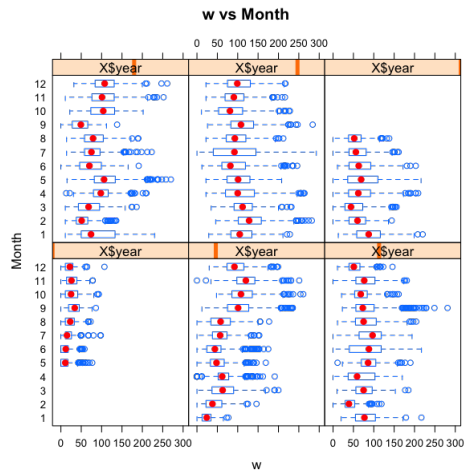
(b) Box plot of sunspot count by year and month.
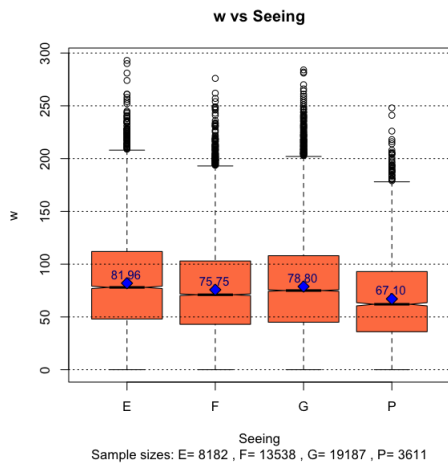
(c) Box plot of sunspot count by seeing condition.
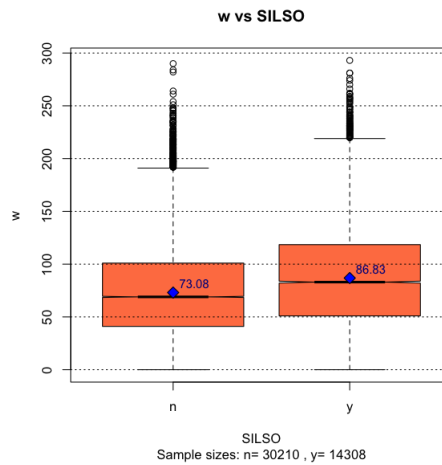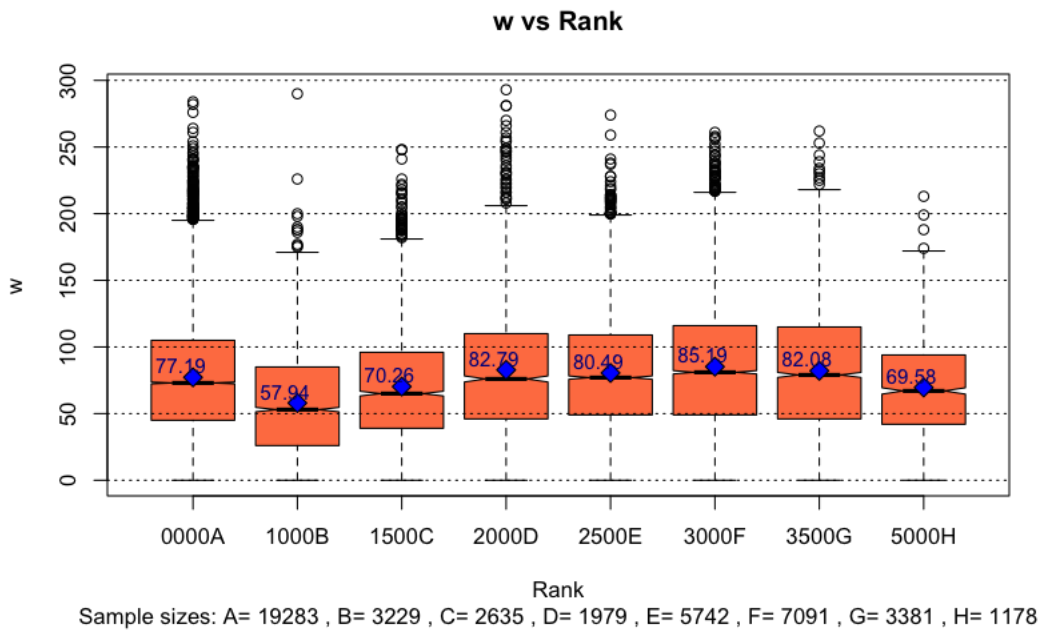

(d) Box plot of sunspot count submitted to AAVSO and SILSO.

Figure 3:

Figure 4: Box plot of sunspot count by rank.