

Monthly Report (00)

2016.02 Data Set

Friday 18th March, 2016

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.

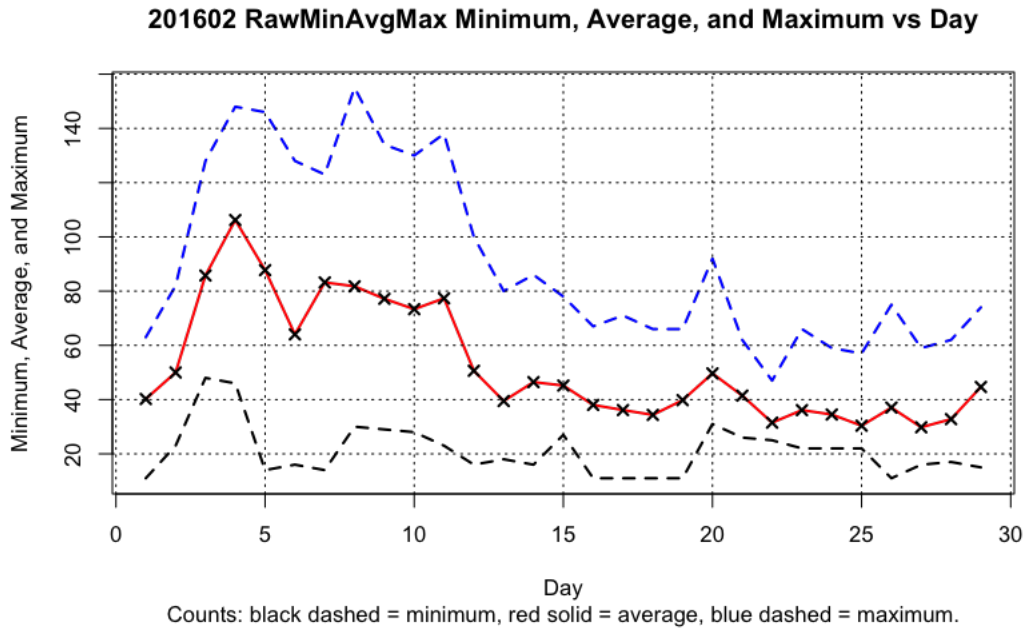


Figure 1: Raw average sunspot count by day of the month.

Table 1: 201602 Daily Raw Counts

Day	Submissions	Minimum	Average	Maximum
1.0000	25.0000	11.0000	40.2222	63.0000
2.0000	24.0000	23.0000	50.0000	82.0000
3.0000	30.0000	48.0000	85.7500	128.0000
4.0000	25.0000	46.0000	106.2083	148.0000
5.0000	27.0000	14.0000	87.7391	146.0000
6.0000	27.0000	16.0000	64.0417	128.0000
7.0000	32.0000	14.0000	83.2069	123.0000
8.0000	33.0000	30.0000	81.7778	155.0000
9.0000	28.0000	29.0000	77.1739	134.0000
10.0000	22.0000	28.0000	73.3500	130.0000
11.0000	26.0000	23.0000	77.3478	138.0000
12.0000	25.0000	16.0000	50.6190	100.0000
13.0000	30.0000	18.0000	39.4783	80.0000
14.0000	27.0000	16.0000	46.4545	86.0000
15.0000	26.0000	27.0000	45.1818	78.0000
16.0000	28.0000	11.0000	38.0385	67.0000
17.0000	37.0000	11.0000	36.1429	71.0000
18.0000	32.0000	11.0000	34.3600	66.0000
19.0000	33.0000	11.0000	39.7917	66.0000
20.0000	28.0000	31.0000	49.6500	92.0000
21.0000	29.0000	26.0000	41.4583	62.0000
22.0000	31.0000	25.0000	31.5000	47.0000
23.0000	28.0000	22.0000	36.1200	66.0000
24.0000	29.0000	22.0000	34.5000	59.0000
25.0000	27.0000	22.0000	30.3750	57.0000
26.0000	31.0000	11.0000	37.0417	75.0000
27.0000	33.0000	16.0000	29.7917	59.0000
28.0000	39.0000	17.0000	32.7931	62.0000
29.0000	33.0000	15.0000	44.6538	74.0000

3 Error Tables

Data are for the month of February 2016. 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	aavso	silso
2010.05	23.3295	22.7964	23.8627	8.4000	8.7000
2010.06	18.7704	18.2684	19.2724	11.0000	13.6000
2010.07	20.5844	20.1306	21.0383	15.2000	16.1000
2010.08	19.6371	19.1588	20.1153	18.3000	19.6000
2010.09	23.6378	23.1311	24.1444	22.8000	25.2000
2010.10	22.6939	22.2037	23.1841	21.0000	23.5000
2010.11	24.0144	23.4699	24.5589	20.9000	21.6000
2010.12	23.2485	22.5797	23.9174	13.9000	14.5000
2011.01	72.6547	71.0215	74.2878	17.7000	18.7000
2011.02	63.5031	62.0639	64.9423	29.1000	29.6000
2011.03	70.9777	69.5105	72.4449	48.0000	55.8000
2011.04	77.7544	76.0845	79.4244	47.3000	54.4000
2011.05	79.2627	77.6697	80.8557	37.3000	41.5000
2011.06	67.2488	65.8295	68.6681	35.2000	37.0000

Continued on next page

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

ym	Ra	lci99	uci99	aavso	silso
2011.07	71.8907	70.3095	73.4718	41.5000	43.8000
2011.08	71.8805	70.4684	73.2926	42.4000	50.5000
2011.09	83.8552	82.7605	84.9498	73.8000	78.0000
2011.10	80.0045	78.6395	81.3695	78.9000	88.0000
2011.11	83.2561	81.4858	85.0263	84.6000	96.7000
2011.12	78.2250	76.5211	79.9289	65.8000	73.0000
2012.01	74.7043	73.2229	76.1856	55.8000	58.2000
2012.02	63.1500	61.8167	64.4834	29.2000	33.1000
2012.03	73.5306	72.2209	74.8403	53.1000	64.1000
2012.04	77.5585	75.3007	79.8163	51.4000	55.2000
2012.05	83.3735	81.9204	84.8266	61.8000	69.0000
2012.06	70.5020	69.2536	71.7504	59.7000	64.5000
2012.07	76.3388	75.0617	77.6159	64.2000	51.3000
2012.08	72.7278	71.5146	73.9410	57.7000	63.1000
2012.09	84.7040	83.2575	86.1505	57.7000	61.5000
2012.10	82.3096	80.7535	83.8656	48.3000	53.3000
2012.11	86.7496	85.0101	88.4890	56.7000	61.4000
2012.12	79.1889	77.5125	80.8652	37.4000	40.8000
2013.01	84.2429	82.6532	85.8326	63.8000	62.9000
2013.02	72.8372	71.4164	74.2580	37.8000	38.0000
2013.03	80.6246	79.1078	82.1413	50.6000	57.9000
2013.04	90.3491	88.8228	91.8754	70.6000	72.4000
2013.05	91.0515	89.4780	92.6250	77.4000	78.7000
2013.06	77.6306	76.2467	79.0146	51.0000	52.5000
2013.07	81.9387	80.6582	83.2192	57.0000	57.0000
2013.08	80.1517	78.8965	81.4069	60.0000	66.0000
2013.09	92.5045	90.9038	94.1053	34.6000	36.9000
2013.10	88.4153	86.8423	89.9883	74.5000	85.6000
2013.11	93.3062	91.3576	95.2547	73.9000	77.6000
2013.12	87.4440	85.6776	89.2104	77.8000	90.3000
2014.01	100.1980	98.0887	102.3072	77.4000	82.0000
2014.02	86.6694	85.0219	88.3168	93.9000	102.8000
2014.03	99.2171	97.5400	100.8942	80.9000	92.2000
2014.04	109.7972	107.9251	111.6693	76.9000	84.7000
2014.05	110.3181	108.5528	112.0834	72.3000	75.2000
2014.06	94.0617	92.5294	95.5941	67.2000	71.0000
2014.07	100.6862	99.0459	102.3265	72.5000	72.5000
2014.08	97.9568	96.4768	99.4369	71.2000	74.7000

Continued on next page

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

ym	Ra	lci99	uci99	aavso	silso
2014.09	114.0502	112.2220	115.8783	83.2000	87.6000
2014.10	108.8512	107.0245	110.6779	59.5000	60.6000
2014.11	115.4159	113.2273	117.6045	65.8000	71.1000
2014.12	105.5290	103.2565	107.8014	75.8000	78.0000
2015.01	60.8665	59.6958	62.0372	65.9000	67.0000
2015.02	52.6316	51.4147	53.8485	42.4000	44.8000
2015.03	59.2040	58.1222	60.2859	38.0000	38.4000
2015.04	66.3484	65.1757	67.5212	49.0000	54.4000
2015.05	66.3684	65.2962	67.4405	56.3000	58.8000
2015.06	56.9731	56.0081	57.9380	50.2000	68.3000
2015.07	59.6477	58.6330	60.6625	47.9000	65.8000
2015.08	59.4020	58.4314	60.3726	39.5000	57.2000
2015.09	69.0044	67.8813	70.1276	49.2000	72.1000
2015.10	65.7873	64.6621	66.9124	39.3000	48.3000
2015.11	70.2981	69.3437	71.2525	39.6000	55.9000
2015.12	63.6888	62.3869	64.9907	36.4000	44.8000
2016.01	53.8210	52.8668	54.7753	33.7000	43.3000
2016.02	45.5289	44.6260	46.4318	38.3000	46.8000

Loglinear Mixed Model Fit, AAVSO, and SILSO Values vs Sequence Boxes and whiskers represent unprocessed counts

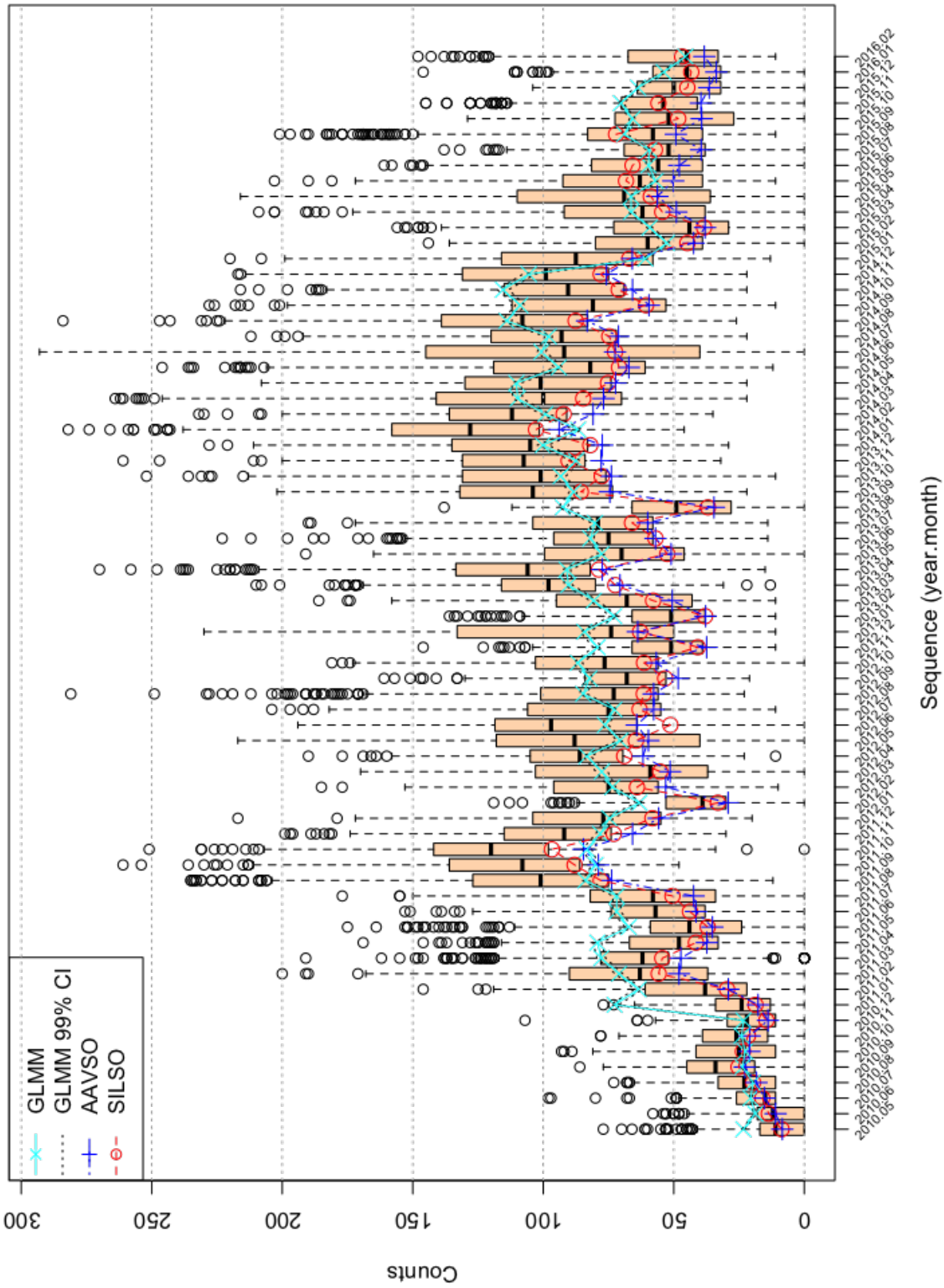


Figure 2: GLMM fitted data for R_a . AAVSO data: <https://www.aavso.org/category/tags/solar-bulletin>. SILSO data: WDC-SILSO, Royal Observatory of Belgium, Brussels

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: 201602 Parameter Estimates

	Estimate	Std. Error	t-value	$\Pr(> t)$
(Intercept)	3.1681	0.0445	71.1539	0.0000
seeF	-0.1848	0.0074	-25.1362	0.0000
seeG	-0.0968	0.0064	-15.1264	0.0000
seeP	-0.2973	0.0108	-27.5882	0.0000
r1000B	-0.0656	0.0831	-0.7895	0.4298
r1500C	0.0275	0.1271	0.2166	0.8285
r2000D	0.0747	0.1550	0.4819	0.6299
r2500E	-0.0005	0.1053	-0.0046	0.9963
r3000F	0.0628	0.1025	0.6130	0.5399
r3500G	0.1182	0.1534	0.7705	0.4410
r5000H	-0.1171	0.2121	-0.5521	0.5809
silsoy	0.1177	0.0739	1.5919	0.1114
year2011	1.2148	0.0152	79.8468	0.0000
year2012	1.2316	0.0152	81.2129	0.0000
year2013	1.3274	0.0151	87.7816	0.0000
year2014	1.5166	0.0150	100.9554	0.0000
year2015	1.0161	0.0154	65.9320	0.0000
year2016	0.8722	0.0238	36.6453	0.0000
mon2	-0.1544	0.0120	-12.9146	0.0000
mon3	-0.0372	0.0114	-3.2553	0.0011
mon4	0.0680	0.0115	5.9238	0.0000
mon5	0.0749	0.0108	6.9005	0.0000
mon6	-0.0936	0.0114	-8.2273	0.0000
mon7	-0.0313	0.0110	-2.8494	0.0044
mon8	-0.0482	0.0109	-4.4333	0.0000
mon9	0.1059	0.0104	10.1774	0.0000
mon10	0.0649	0.0110	5.9071	0.0000
mon11	0.1276	0.0112	11.4121	0.0000
mon12	0.0490	0.0119	4.1246	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.

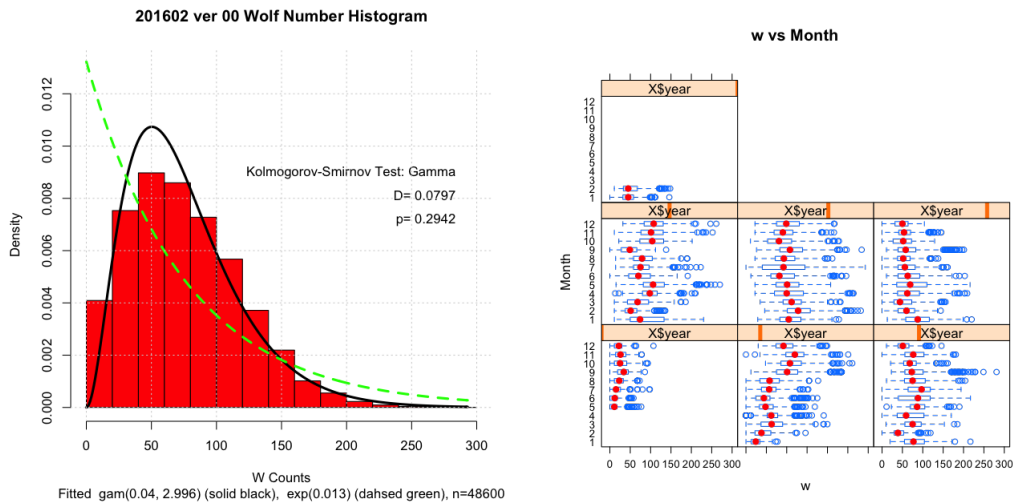
5 Supporting Information

Table 4: 201602 Summary of Sunspot Numbers

obs	jd	year	mon	day
ARAG : 2077	Min. :1721096	Min. :2010	Min. : 1.000	Min. : 1.00
CHAG : 1897	1st Qu.:2455919	1st Qu.:2011	1st Qu.: 4.000	1st Qu.: 8.00
BRAB : 1866	Median :2456428	Median :2013	Median : 7.000	Median :16.00
BROB : 1671	Mean :2456056	Mean :2013	Mean : 6.699	Mean :15.73
DUBF : 1564	3rd Qu.:2456905	3rd Qu.:2014	3rd Qu.: 9.000	3rd Qu.:23.00
HOWR : 1543	Max. :2457448	Max. :2016	Max. :12.000	Max. :31.00
(Other):37982				

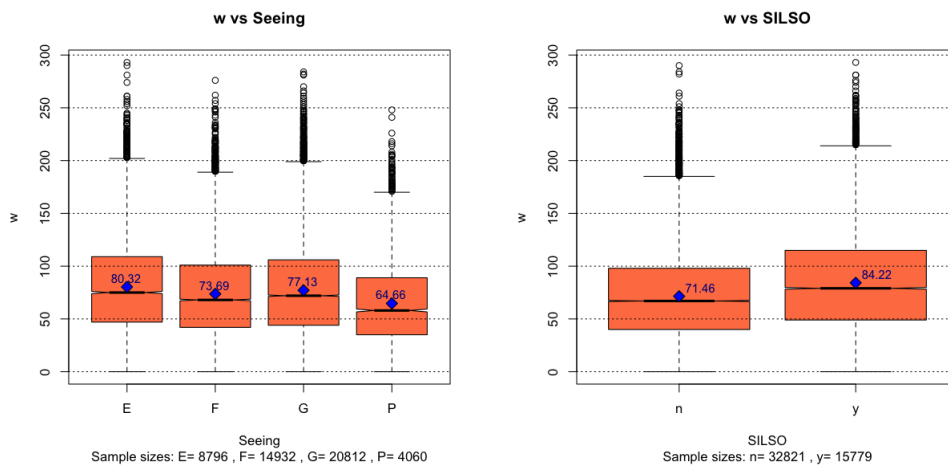
Table 5: Summary of Sunspot Numbers

see	g	s	w	r	silso
E: 8796	Min. : 0.000	Min. : 0.00	Min. : 0.00	0000A :20799	n:32821
F:14932	1st Qu.: 3.000	1st Qu.: 12.00	1st Qu.: 43.00	3000F : 7895	y:15779
G:20812	Median : 4.000	Median : 24.00	Median : 70.00	2500E : 6296	
P: 4060	Mean : 4.633	Mean : 29.27	Mean : 75.61	3500G : 3763	
	3rd Qu.: 6.000	3rd Qu.: 42.00	3rd Qu.:104.00	1000B : 3520	
	Max. :18.000	Max. :204.00	Max. :293.00	1500C : 2871	
				(Other): 3456	



(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:

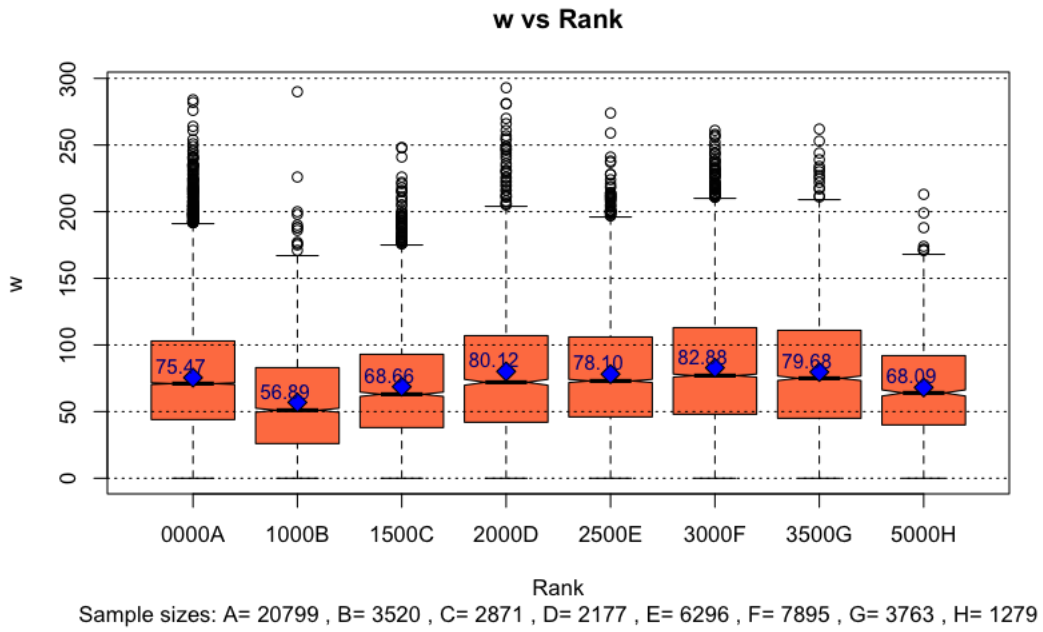


Figure 4: Box plot of sunspot count by rank.