

Monthly Report (00)

2016.04 Data Set

Monday 16th May, 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.

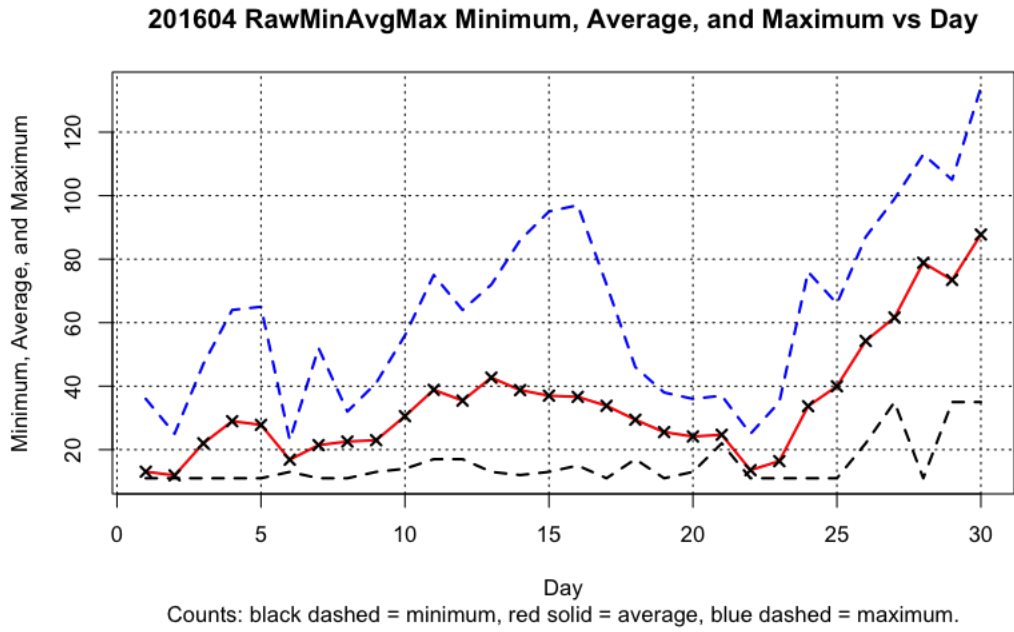


Figure 1: Raw average sunspot count by day of the month.

Table 1: 201604 Daily Raw Counts

Day	Submissions	Minimum	Average	Maximum
1.0000	35.0000	11.0000	13.0606	36.0000
2.0000	33.0000	11.0000	11.9000	25.0000
3.0000	38.0000	11.0000	22.0000	47.0000
4.0000	32.0000	11.0000	28.9655	64.0000
5.0000	36.0000	11.0000	27.8182	65.0000
6.0000	32.0000	13.0000	16.8333	23.0000
7.0000	27.0000	11.0000	21.4444	52.0000
8.0000	27.0000	11.0000	22.5556	32.0000
9.0000	35.0000	13.0000	22.9688	41.0000
10.0000	34.0000	14.0000	30.5517	56.0000
11.0000	33.0000	17.0000	38.8438	75.0000
12.0000	32.0000	17.0000	35.4516	64.0000
13.0000	34.0000	13.0000	42.6897	72.0000
14.0000	37.0000	12.0000	38.7576	86.0000
15.0000	35.0000	13.0000	37.0000	95.0000
16.0000	33.0000	15.0000	36.6333	97.0000
17.0000	37.0000	11.0000	33.8125	72.0000
18.0000	39.0000	17.0000	29.4242	46.0000
19.0000	39.0000	11.0000	25.5294	38.0000
20.0000	39.0000	13.0000	24.1471	36.0000
21.0000	38.0000	22.0000	24.6970	37.0000
22.0000	34.0000	11.0000	13.5484	25.0000
23.0000	35.0000	11.0000	16.3939	35.0000
24.0000	37.0000	11.0000	33.7059	76.0000
25.0000	36.0000	11.0000	39.9706	66.0000
26.0000	28.0000	22.0000	54.2500	87.0000
27.0000	31.0000	35.0000	61.6296	99.0000
28.0000	33.0000	11.0000	78.8667	113.0000
29.0000	28.0000	35.0000	73.4444	105.0000
30.0000	29.0000	35.0000	87.7600	134.0000

3 Error Tables

Data are for the month of April 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	23.3274	22.7943	23.8604	8.4000	8.7000
2010.06	18.7654	18.2633	19.2676	11.0000	13.6000
2010.07	20.5843	20.1306	21.0380	15.2000	16.1000
2010.08	19.6328	19.1541	20.1115	18.3000	19.6000
2010.09	23.6392	23.1322	24.1462	22.8000	25.2000
2010.10	22.7021	22.2118	23.1924	21.0000	23.5000
2010.11	24.0176	23.4732	24.5619	20.9000	21.6000
2010.12	23.2451	22.5759	23.9143	13.9000	14.5000
2011.01	74.2315	72.5642	75.8989	17.7000	18.7000
2011.02	64.9015	63.4322	66.3709	29.1000	29.6000
2011.03	70.5567	69.0988	72.0145	48.0000	55.8000
2011.04	75.6540	74.0306	77.2774	47.3000	54.4000
2011.05	79.2805	77.6886	80.8723	37.3000	41.5000
2011.06	67.2683	65.8503	68.6863	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	71.9136	70.3337	73.4934	41.5000	43.8000
2011.08	71.9023	70.4910	73.3136	42.4000	50.5000
2011.09	83.8553	82.7587	84.9519	73.8000	78.0000
2011.10	80.0084	78.6405	81.3763	78.9000	88.0000
2011.11	83.2353	81.4602	85.0103	84.6000	96.7000
2011.12	78.2139	76.5060	79.9219	65.8000	73.0000
2012.01	76.1776	74.6664	77.6888	55.8000	58.2000
2012.02	64.4193	63.0564	65.7821	29.2000	33.1000
2012.03	72.9546	71.6534	74.2559	53.1000	64.1000
2012.04	75.3288	73.1319	77.5258	51.4000	55.2000
2012.05	83.2438	81.7915	84.6962	61.8000	69.0000
2012.06	70.3957	69.1480	71.6434	59.7000	64.5000
2012.07	76.2190	74.9427	77.4952	64.2000	51.3000
2012.08	72.6043	71.3915	73.8170	57.7000	63.1000
2012.09	84.5741	83.1291	86.0192	57.7000	61.5000
2012.10	82.1835	80.6287	83.7382	48.3000	53.3000
2012.11	86.6117	84.8735	88.3500	56.7000	61.4000
2012.12	79.0632	77.3876	80.7387	37.4000	40.8000
2013.01	86.0868	84.4620	87.7116	63.8000	62.9000
2013.02	74.4866	73.0315	75.9418	37.8000	38.0000
2013.03	80.1863	78.6757	81.6969	50.6000	57.9000
2013.04	87.9277	86.4409	89.4144	70.6000	72.4000
2013.05	91.1014	89.5256	92.6772	77.4000	78.7000
2013.06	77.6726	76.2863	79.0590	51.0000	52.5000
2013.07	81.9793	80.6970	83.2615	57.0000	57.0000
2013.08	80.1944	78.9376	81.4512	60.0000	66.0000
2013.09	92.5398	90.9359	94.1437	34.6000	36.9000
2013.10	88.4447	86.8689	90.0204	74.5000	85.6000
2013.11	93.3509	91.3989	95.3030	73.9000	77.6000
2013.12	87.4840	85.7158	89.2521	77.8000	90.3000
2014.01	102.3659	100.2068	104.5250	77.4000	82.0000
2014.02	88.5755	86.8873	90.2636	93.9000	102.8000
2014.03	98.6673	96.9998	100.3349	80.9000	92.2000
2014.04	106.8461	105.0230	108.6692	76.9000	84.7000
2014.05	110.3591	108.5912	112.1270	72.3000	75.2000
2014.06	94.0891	92.5549	95.6233	67.2000	71.0000
2014.07	100.7267	99.0844	102.3690	72.5000	72.5000
2014.08	98.0084	96.5264	99.4903	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	114.1071	112.2758	115.9384	83.2000	87.6000
2014.10	108.8968	107.0672	110.7264	59.5000	60.6000
2014.11	115.4602	113.2662	117.6543	65.8000	71.1000
2014.12	105.5878	103.3091	107.8664	75.8000	78.0000
2015.01	62.1931	60.9938	63.3924	65.9000	67.0000
2015.02	53.8192	52.5702	55.0682	42.4000	44.8000
2015.03	58.8883	57.8109	59.9658	38.0000	38.4000
2015.04	64.5829	63.4400	65.7259	49.0000	54.4000
2015.05	66.4040	65.3308	67.4772	56.3000	58.8000
2015.06	57.0091	56.0428	57.9753	50.2000	68.3000
2015.07	59.6804	58.6642	60.6967	47.9000	65.8000
2015.08	59.4324	58.4601	60.4047	39.5000	57.2000
2015.09	69.0554	67.9317	70.1791	49.2000	72.1000
2015.10	65.8308	64.7050	66.9566	39.3000	48.3000
2015.11	70.3446	69.3895	71.2997	39.6000	55.9000
2015.12	63.7439	62.4431	65.0447	36.4000	44.8000
2016.01	44.6062	43.8149	45.3975	33.7000	43.3000
2016.02	37.7659	37.0162	38.5156	38.3000	46.8000
2016.03	41.1665	40.4135	41.9194	30.5000	38.9000
2016.04	44.3908	43.6011	45.1805	26.6000	30.9000

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

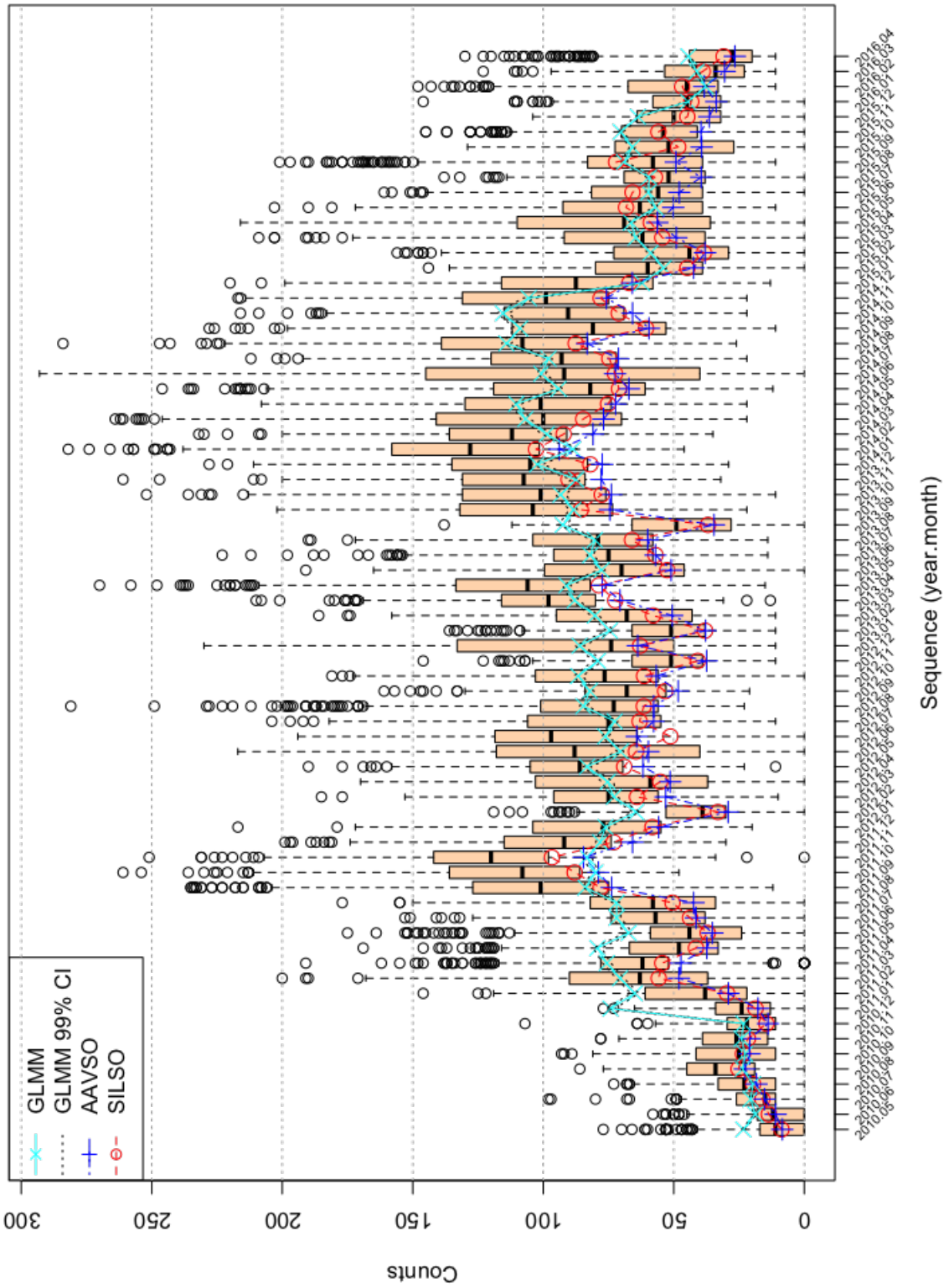


Figure 2: GLMM fitted data for R_a . AAVSO data: <https://www.aavso.org/category/tags/solar-bulletin>. SILSO data: WDC-SILSO, Royal Observatory of Belgium, Brussels

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: 201604 Parameter Estimates

	Estimate	Std. Error	t-value	$\Pr(> t)$
(Intercept)	3.1885	0.0445	71.6203	0.0000
seeF	-0.1848	0.0073	-25.3596	0.0000
seeG	-0.0975	0.0063	-15.3708	0.0000
seeP	-0.2962	0.0107	-27.7474	0.0000
r1000B	-0.0653	0.0832	-0.7852	0.4324
r1500C	0.0283	0.1272	0.2224	0.8240
r2000D	0.0749	0.1552	0.4827	0.6293
r2500E	-0.0000	0.1054	-0.0004	0.9997
r3000F	0.0648	0.1027	0.6317	0.5276
r3500G	0.1191	0.1536	0.7751	0.4383
r5000H	-0.1155	0.2124	-0.5438	0.5866
silsoy	0.1172	0.0740	1.5839	0.1132
year2011	1.2151	0.0152	80.0826	0.0000
year2012	1.2302	0.0151	81.3471	0.0000
year2013	1.3283	0.0151	88.0779	0.0000
year2014	1.5174	0.0150	101.2808	0.0000
year2015	1.0169	0.0154	66.1676	0.0000
year2016	0.6634	0.0197	33.6051	0.0000
mon2	-0.1537	0.0119	-12.8948	0.0000
mon3	-0.0641	0.0110	-5.8500	0.0000
mon4	0.0194	0.0110	1.7621	0.0781
mon5	0.0540	0.0107	5.0474	0.0000
mon6	-0.1144	0.0112	-10.1972	0.0000
mon7	-0.0521	0.0108	-4.8211	0.0000
mon8	-0.0690	0.0107	-6.4453	0.0000
mon9	0.0851	0.0102	8.3092	0.0000
mon10	0.0439	0.0108	4.0581	0.0000
mon11	0.1065	0.0110	9.6645	0.0000
mon12	0.0278	0.0117	2.3761	0.0175

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.

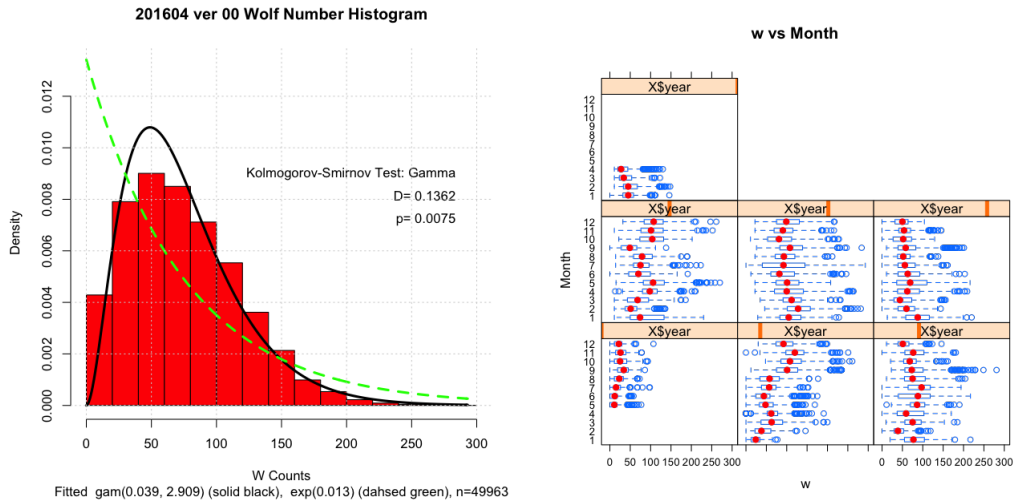
5 Supporting Information

Table 4: 201604 Summary of Sunspot Numbers

obs	jd	year	mon	day
ARAG : 2138	Min. :1721096	Min. :2010	Min. : 1.000	Min. : 1.00
CHAG : 1957	1st Qu.:2455935	1st Qu.:2012	1st Qu.: 4.000	1st Qu.: 8.00
BRAB : 1924	Median :2456453	Median :2013	Median : 7.000	Median :16.00
BROB : 1720	Mean :2456094	Mean :2013	Mean : 6.612	Mean :15.73
DUBF : 1619	3rd Qu.:2456945	3rd Qu.:2014	3rd Qu.: 9.000	3rd Qu.:23.00
HOWR : 1592	Max. :2457509	Max. :2016	Max. :12.000	Max. :31.00
(Other):39013				

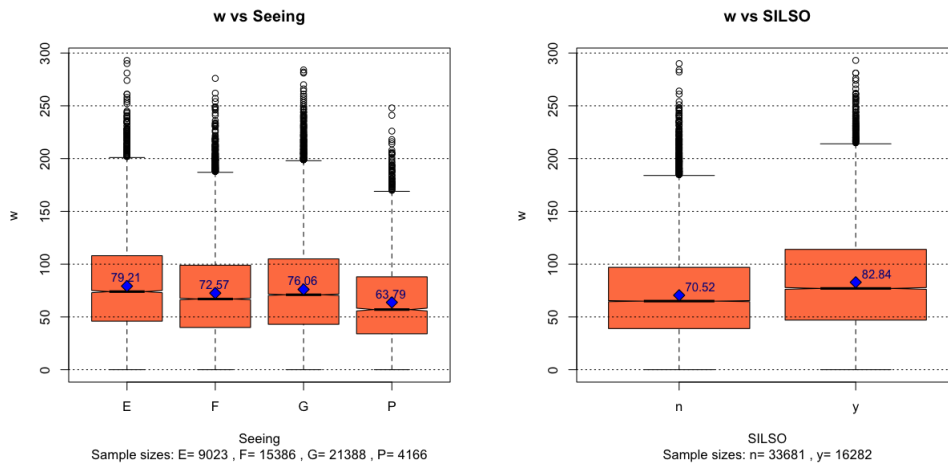
Table 5: Summary of Sunspot Numbers

see	g	s	w	r	silso
E: 9023	Min. : 0.000	Min. : 0.00	Min. : 0.00	0000A :21387	n:33681
F:15386	1st Qu.: 3.000	1st Qu.: 11.00	1st Qu.: 41.00	3000F : 8137	y:16282
G:21388	Median : 4.000	Median : 23.00	Median : 69.00	2500E : 6474	
P: 4166	Mean : 4.578	Mean : 28.75	Mean : 74.53	3500G : 3881	
	3rd Qu.: 6.000	3rd Qu.: 41.00	3rd Qu.:103.00	1000B : 3596	
	Max. :18.000	Max. :204.00	Max. :293.00	1500C : 2919	
				(Other): 3569	



(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:

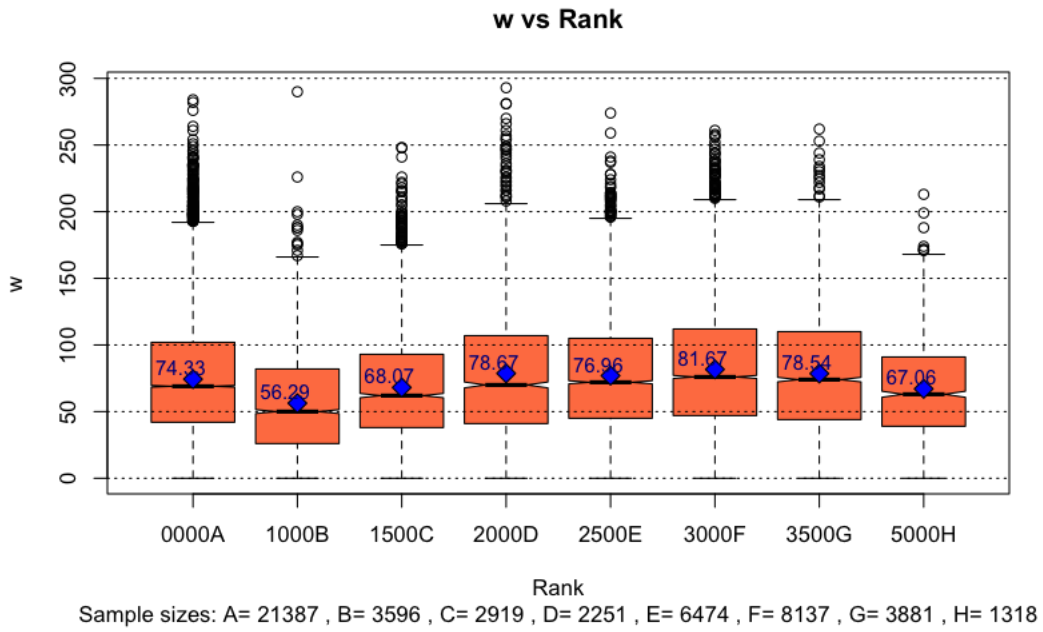


Figure 4: Box plot of sunspot count by rank.