

# Monthly Report (00)

## 2016.01 Data Set

Thursday 18<sup>th</sup> February, 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.

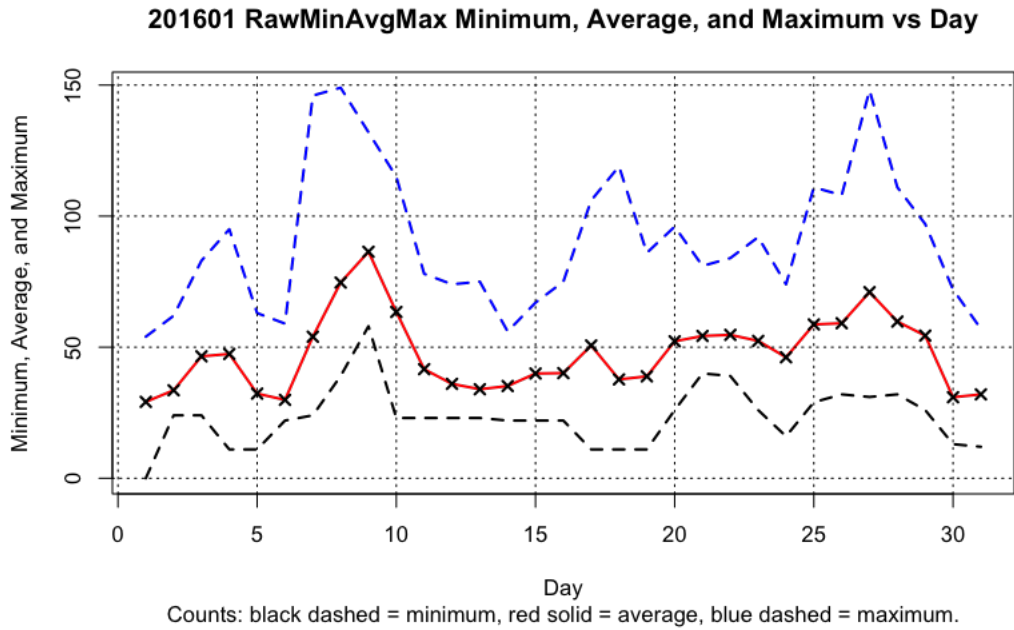


Figure 1: Raw average sunspot count by day of the month.

Table 1: 201601 Daily Raw Counts

Day	Submissions	Minimum	Average	Maximum
1.0000	24.0000	0.0000	29.1429	54.0000
2.0000	31.0000	24.0000	33.5652	62.0000
3.0000	24.0000	24.0000	46.5238	83.0000
4.0000	23.0000	11.0000	47.4211	95.0000
5.0000	29.0000	11.0000	32.2917	63.0000
6.0000	23.0000	22.0000	29.8947	59.0000
7.0000	21.0000	24.0000	54.0000	146.0000
8.0000	25.0000	39.0000	74.6667	149.0000
9.0000	25.0000	58.0000	86.4091	132.0000
10.0000	28.0000	23.0000	63.4400	115.0000
11.0000	29.0000	23.0000	41.6154	78.0000
12.0000	33.0000	23.0000	36.0000	74.0000
13.0000	27.0000	23.0000	33.9600	75.0000
14.0000	24.0000	22.0000	35.1818	56.0000
15.0000	30.0000	22.0000	39.9600	67.0000
16.0000	30.0000	22.0000	40.1200	75.0000
17.0000	29.0000	11.0000	50.6522	106.0000
18.0000	29.0000	11.0000	37.7083	119.0000
19.0000	29.0000	11.0000	38.8800	86.0000
20.0000	24.0000	26.0000	52.2500	96.0000
21.0000	31.0000	40.0000	54.2800	81.0000
22.0000	28.0000	39.0000	54.7083	84.0000
23.0000	33.0000	26.0000	52.4000	92.0000
24.0000	32.0000	16.0000	46.1538	74.0000
25.0000	32.0000	29.0000	58.6800	111.0000
26.0000	25.0000	32.0000	59.1429	108.0000
27.0000	24.0000	31.0000	71.0000	148.0000
28.0000	32.0000	32.0000	59.7778	111.0000
29.0000	26.0000	26.0000	54.4167	97.0000
30.0000	25.0000	13.0000	30.9130	72.0000
31.0000	25.0000	12.0000	32.0000	57.0000

### 3 Error Tables

Data are for the month of January 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 25<sup>th</sup> through the 75<sup>th</sup> quartiles. The lower and upper whiskers extend 1.5 times the IQR below the 25<sup>th</sup> quartile, and 1.5 times the IQR above the 75<sup>th</sup> 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	23.3307	22.7977	23.8637	8.4000	8.7000
2010.06	18.7729	18.2712	19.2746	11.0000	13.6000
2010.07	20.5862	20.1323	21.0402	15.2000	16.1000
2010.08	19.6384	19.1603	20.1166	18.3000	19.6000
2010.09	23.6390	23.1325	24.1455	22.8000	25.2000
2010.10	22.6873	22.1965	23.1781	21.0000	23.5000
2010.11	24.0114	23.4668	24.5559	20.9000	21.6000
2010.12	23.2511	22.5816	23.9207	13.9000	14.5000
2011.01	73.8090	72.1477	75.4703	17.7000	18.7000
2011.02	62.2511	60.8415	63.6607	29.1000	29.6000
2011.03	70.9734	69.5064	72.4404	48.0000	55.8000
2011.04	77.7444	76.0735	79.4153	47.3000	54.4000
2011.05	79.2631	77.6695	80.8567	37.3000	41.5000
2011.06	67.2496	65.8297	68.6696	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	71.8816	70.2986	73.4646	41.5000	43.8000
2011.08	71.8654	70.4519	73.2790	42.4000	50.5000
2011.09	83.8524	82.7580	84.9468	73.8000	78.0000
2011.10	80.0062	78.6411	81.3713	78.9000	88.0000
2011.11	83.2630	81.4937	85.0324	84.6000	96.7000
2011.12	78.2281	76.5254	79.9308	65.8000	73.0000
2012.01	75.9174	74.4113	77.4235	55.8000	58.2000
2012.02	61.9263	60.6194	63.2332	29.2000	33.1000
2012.03	73.5450	72.2360	74.8541	53.1000	64.1000
2012.04	77.5345	75.2797	79.7893	51.4000	55.2000
2012.05	83.3742	81.9214	84.8269	61.8000	69.0000
2012.06	70.4993	69.2509	71.7477	59.7000	64.5000
2012.07	76.3404	75.0628	77.6180	64.2000	51.3000
2012.08	72.7419	71.5294	73.9543	57.7000	63.1000
2012.09	84.7056	83.2599	86.1513	57.7000	61.5000
2012.10	82.3136	80.7588	83.8685	48.3000	53.3000
2012.11	86.7567	85.0180	88.4955	56.7000	61.4000
2012.12	79.1910	77.5163	80.8656	37.4000	40.8000
2013.01	85.6035	83.9894	87.2176	63.8000	62.9000
2013.02	71.3883	69.9968	72.7798	37.8000	38.0000
2013.03	80.6113	79.0958	82.1268	50.6000	57.9000
2013.04	90.3554	88.8307	91.8800	70.6000	72.4000
2013.05	91.0543	89.4828	92.6258	77.4000	78.7000
2013.06	77.6362	76.2536	79.0189	51.0000	52.5000
2013.07	81.9395	80.6607	83.2183	57.0000	57.0000
2013.08	80.1471	78.8926	81.4016	60.0000	66.0000
2013.09	92.5034	90.9038	94.1030	34.6000	36.9000
2013.10	88.4259	86.8535	89.9984	74.5000	85.6000
2013.11	93.3185	91.3719	95.2651	73.9000	77.6000
2013.12	87.4438	85.6791	89.2085	77.8000	90.3000
2014.01	101.8282	99.6865	103.9700	77.4000	82.0000
2014.02	84.9819	83.3685	86.5953	93.9000	102.8000
2014.03	99.2222	97.5453	100.8991	80.9000	92.2000
2014.04	109.8158	107.9454	111.6861	76.9000	84.7000
2014.05	110.3210	108.5566	112.0854	72.3000	75.2000
2014.06	94.0691	92.5386	95.5996	67.2000	71.0000
2014.07	100.6937	99.0545	102.3328	72.5000	72.5000
2014.08	97.9659	96.4870	99.4448	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	114.0664	112.2402	115.8926	83.2000	87.6000
2014.10	108.8565	107.0304	110.6826	59.5000	60.6000
2014.11	115.4269	113.2397	117.6142	65.8000	71.1000
2014.12	105.5410	103.2699	107.8121	75.8000	78.0000
2015.01	61.8429	60.6546	63.0311	65.9000	67.0000
2015.02	51.5962	50.4058	52.7865	42.4000	44.8000
2015.03	59.1979	58.1178	60.2781	38.0000	38.4000
2015.04	66.3372	65.1662	67.5081	49.0000	54.4000
2015.05	66.3604	65.2898	67.4311	56.3000	58.8000
2015.06	56.9597	55.9960	57.9233	50.2000	68.3000
2015.07	59.6425	58.6291	60.6559	47.9000	66.4000
2015.08	59.3903	58.4206	60.3600	39.5000	64.6000
2015.09	68.9922	67.8702	70.1141	49.2000	78.1000
2015.10	65.7692	64.6451	66.8932	39.3000	61.7000
2015.11	70.2793	69.3256	71.2330	39.6000	63.2000
2015.12	63.6706	62.3706	64.9707	36.4000	57.7000
2016.01	46.8518	46.0219	47.6817	33.4000	56.6000

---



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

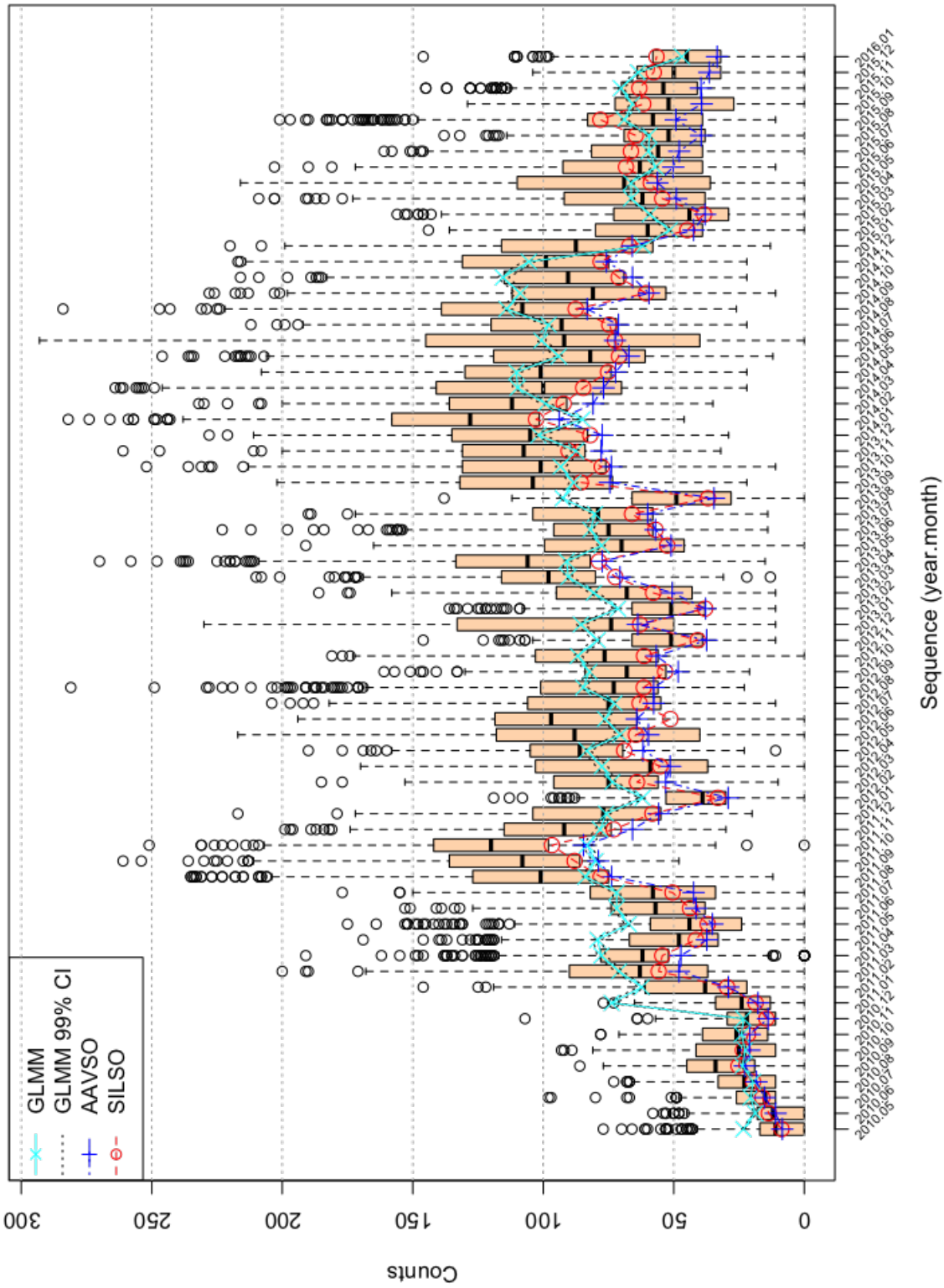


Figure 2: GLMM fitted data for  $R_a$ . AAVSO data: <https://www.aavso.org/category/tags/solar-bulletin>. SILSO data: WDC-SILSO, Royal Observatory of Belgium, Brussels

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: 201601 Parameter Estimates

	Estimate	Std. Error	t-value	$\Pr(> t )$
(Intercept)	3.1851	0.0446	71.4833	0.0000
seeF	-0.1854	0.0074	-25.1112	0.0000
seeG	-0.0978	0.0064	-15.2117	0.0000
seeP	-0.2976	0.0108	-27.4532	0.0000
r1000B	-0.0657	0.0831	-0.7903	0.4294
r1500C	0.0271	0.1270	0.2134	0.8311
r2000D	0.0740	0.1550	0.4772	0.6332
r2500E	-0.0011	0.1053	-0.0103	0.9918
r3000F	0.0615	0.1025	0.6001	0.5484
r3500G	0.1173	0.1534	0.7647	0.4445
r5000H	-0.1179	0.2121	-0.5556	0.5785
silsoy	0.1180	0.0739	1.5970	0.1103
year2011	1.2148	0.0152	79.7557	0.0000
year2012	1.2317	0.0152	81.1253	0.0000
year2013	1.3274	0.0151	87.6761	0.0000
year2014	1.5167	0.0150	100.8399	0.0000
year2015	1.0158	0.0154	65.8373	0.0000
year2016	0.7174	0.0309	23.2447	0.0000
mon2	-0.1904	0.0127	-14.9622	0.0000
mon3	-0.0532	0.0115	-4.6146	0.0000
mon4	0.0519	0.0116	4.4713	0.0000
mon5	0.0588	0.0110	5.3524	0.0000
mon6	-0.1097	0.0115	-9.5407	0.0000
mon7	-0.0474	0.0111	-4.2696	0.0000
mon8	-0.0643	0.0110	-5.8488	0.0000
mon9	0.0898	0.0105	8.5197	0.0000
mon10	0.0488	0.0111	4.3925	0.0000
mon11	0.1115	0.0113	9.8673	0.0000
mon12	0.0329	0.0120	2.7422	0.0061

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.

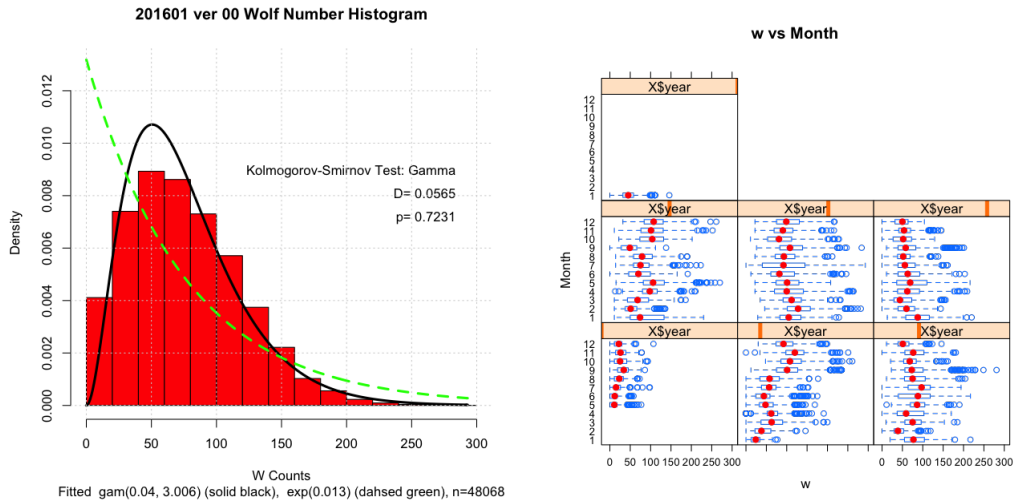
## 5 Supporting Information

Table 4: 201601 Summary of Sunspot Numbers

obs	jd	year	mon	day
ARAG : 2049	Min. :1721096	Min. :2010	Min. : 1.000	Min. : 1.00
CHAG : 1873	1st Qu.:2455912	1st Qu.:2011	1st Qu.: 4.000	1st Qu.: 8.00
BRAB : 1838	Median :2456418	Median :2013	Median : 7.000	Median :16.00
BROB : 1645	Mean :2456040	Mean :2013	Mean : 6.751	Mean :15.73
DUBF : 1544	3rd Qu.:2456892	3rd Qu.:2014	3rd Qu.: 9.000	3rd Qu.:23.00
HOWR : 1517	Max. :2457419	Max. :2016	Max. :12.000	Max. :31.00
(Other):37602				

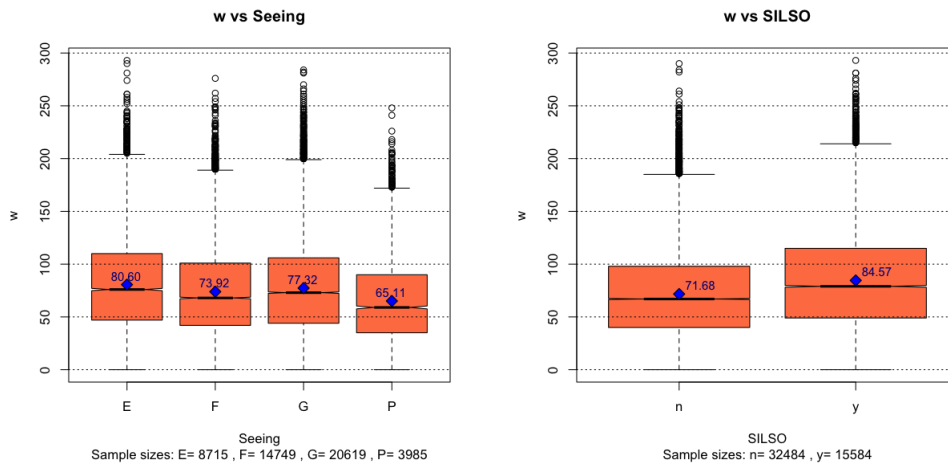
Table 5: Summary of Sunspot Numbers

see	g	s	w	r	silso
E: 8715	Min. : 0.000	Min. : 0.00	Min. : 0.00	0000A :20618	n:32484
F:14749	1st Qu.: 3.000	1st Qu.: 12.00	1st Qu.: 43.00	3000F : 7769	y:15584
G:20619	Median : 4.000	Median : 24.00	Median : 71.00	2500E : 6231	
P: 3985	Mean : 4.649	Mean : 29.37	Mean : 75.86	3500G : 3711	
	3rd Qu.: 6.000	3rd Qu.: 42.00	3rd Qu.:104.00	1000B : 3487	
	Max. :18.000	Max. :204.00	Max. :293.00	1500C : 2845	
				(Other): 3407	



(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:

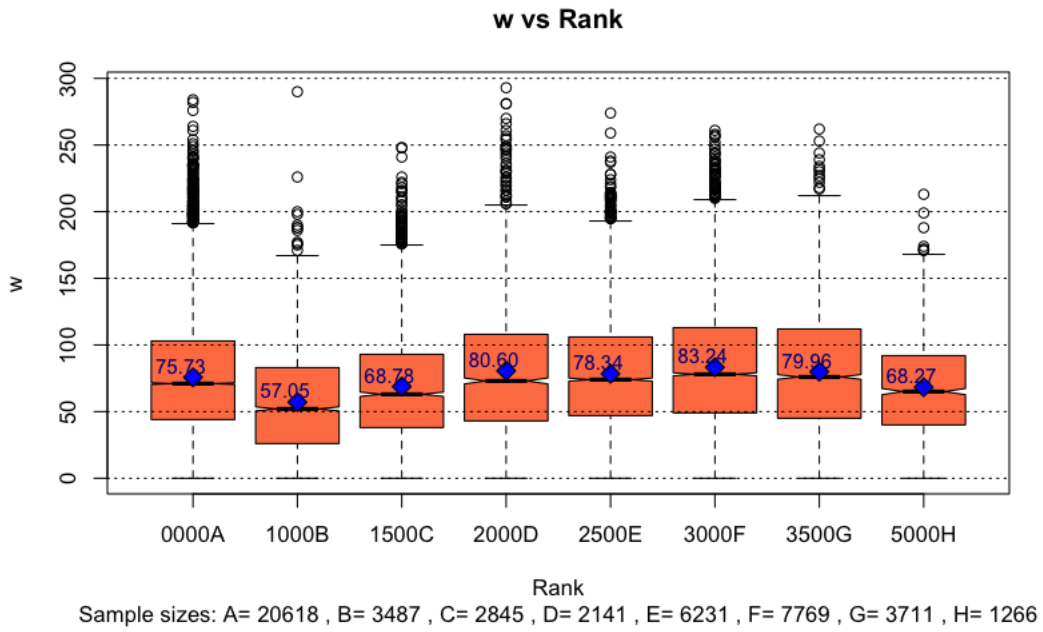


Figure 4: Box plot of sunspot count by rank.