

Monthly Report (00)

2016.08 Data Set

Thursday 15th September, 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.

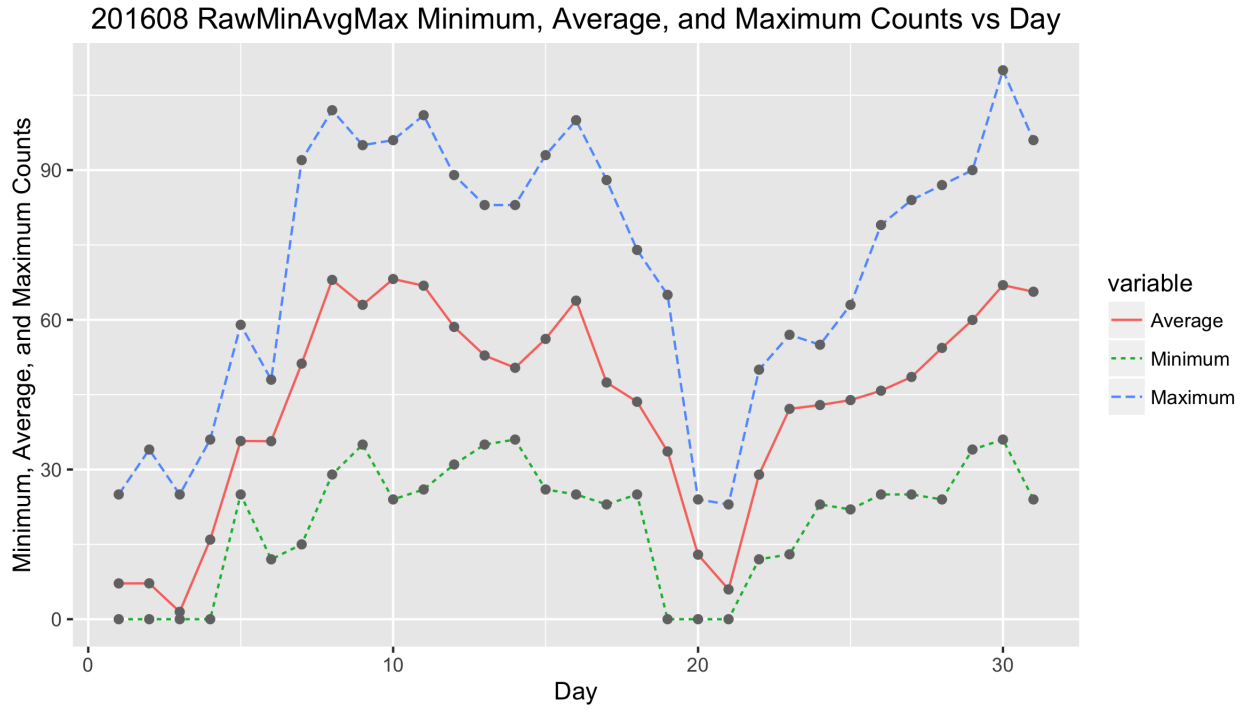


Figure 1: Raw average sunspot count by day of the month.

Table 1: 201608 Daily Raw Counts

Day	Submissions	Minimum	Average	Maximum
1.0000	35.0000	0.0000	7.1765	25.0000
2.0000	32.0000	0.0000	7.1875	34.0000
3.0000	33.0000	0.0000	1.4848	25.0000
4.0000	34.0000	0.0000	15.9412	36.0000
5.0000	31.0000	25.0000	35.7097	59.0000
6.0000	39.0000	12.0000	35.6757	48.0000
7.0000	40.0000	15.0000	51.2162	92.0000
8.0000	41.0000	29.0000	68.0000	102.0000
9.0000	32.0000	35.0000	63.0000	95.0000
10.0000	37.0000	24.0000	68.1714	96.0000
11.0000	37.0000	26.0000	66.8333	101.0000
12.0000	34.0000	31.0000	58.5758	89.0000
13.0000	34.0000	35.0000	52.8485	83.0000
14.0000	35.0000	36.0000	50.3824	83.0000
15.0000	35.0000	26.0000	56.1667	93.0000
16.0000	34.0000	25.0000	63.8387	100.0000
17.0000	37.0000	23.0000	47.4412	88.0000
18.0000	38.0000	25.0000	43.5714	74.0000
19.0000	35.0000	0.0000	33.6176	65.0000
20.0000	34.0000	0.0000	12.9355	24.0000
21.0000	32.0000	0.0000	5.9667	23.0000
22.0000	43.0000	12.0000	28.9737	50.0000
23.0000	43.0000	13.0000	42.1500	57.0000
24.0000	40.0000	23.0000	42.9189	55.0000
25.0000	33.0000	22.0000	43.9062	63.0000
26.0000	41.0000	25.0000	45.7895	79.0000
27.0000	37.0000	25.0000	48.5429	84.0000
28.0000	40.0000	24.0000	54.3714	87.0000
29.0000	34.0000	34.0000	59.9697	90.0000
30.0000	40.0000	36.0000	66.9412	110.0000
31.0000	36.0000	24.0000	65.6250	96.0000

3 Error Tables

Data are for the month of August 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	23.5557	23.0182	24.0932	8.4000	8.7000
2010.06	18.0310	17.5497	18.5123	11.0000	13.6000
2010.07	20.2963	19.8485	20.7441	15.2000	16.1000
2010.08	20.0204	19.5329	20.5079	18.3000	19.6000
2010.09	23.7417	23.2330	24.2505	22.8000	25.2000
2010.10	22.8005	22.3084	23.2927	21.0000	23.5000
2010.11	24.1202	23.5753	24.6651	20.9000	21.6000
2010.12	23.3415	22.6701	24.0128	13.9000	14.5000
2011.01	74.9751	73.2921	76.6581	17.7000	18.7000
2011.02	65.5696	64.0913	67.0479	29.1000	29.6000
2011.03	71.2394	69.7691	72.7097	48.0000	55.8000
2011.04	76.5810	74.9423	78.2196	47.3000	54.4000
2011.05	79.8034	78.2103	81.3964	37.3000	41.5000
2011.06	64.3602	63.0073	65.7132	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	70.6071	69.0565	72.1576	41.5000	43.8000
2011.08	72.9768	71.5484	74.4052	42.4000	50.5000
2011.09	83.7765	82.6767	84.8763	73.8000	78.0000
2011.10	79.9140	78.5442	81.2839	78.9000	88.0000
2011.11	83.1332	81.3543	84.9121	84.6000	96.7000
2011.12	78.0813	76.3726	79.7900	65.8000	73.0000
2012.01	77.0506	75.5222	78.5789	55.8000	58.2000
2012.02	65.1783	63.7970	66.5597	29.2000	33.1000
2012.03	73.7420	72.4244	75.0595	53.1000	64.1000
2012.04	76.2993	74.0503	78.5482	51.4000	55.2000
2012.05	83.8075	82.3430	85.2721	61.8000	69.0000
2012.06	67.4126	66.2145	68.6107	59.7000	64.5000
2012.07	74.9084	73.6471	76.1696	64.2000	51.3000
2012.08	73.7683	72.5353	75.0013	57.7000	63.1000
2012.09	84.6187	83.1733	86.0642	57.7000	61.5000
2012.10	82.2307	80.6774	83.7840	48.3000	53.3000
2012.11	86.6810	84.9430	88.4190	56.7000	61.4000
2012.12	79.1353	77.4656	80.8049	37.4000	40.8000
2013.01	86.9895	85.3547	88.6243	63.8000	62.9000
2013.02	75.3042	73.8367	76.7718	37.8000	38.0000
2013.03	81.0463	79.5214	82.5712	50.6000	57.9000
2013.04	89.0346	87.5337	90.5355	70.6000	72.4000
2013.05	91.6281	90.0453	93.2110	77.4000	78.7000
2013.06	74.3195	72.9946	75.6443	51.0000	52.5000
2013.07	80.5192	79.2659	81.7726	57.0000	57.0000
2013.08	81.4091	80.1375	82.6807	60.0000	66.0000
2013.09	92.5324	90.9324	94.1324	34.6000	36.9000
2013.10	88.4322	86.8585	90.0060	74.5000	85.6000
2013.11	93.3079	91.3632	95.2526	73.9000	77.6000
2013.12	87.4174	85.6578	89.1771	77.8000	90.3000
2014.01	103.4487	101.2730	105.6245	77.4000	82.0000
2014.02	89.5172	87.8123	91.2221	93.9000	102.8000
2014.03	99.6812	97.9992	101.3633	80.9000	92.2000
2014.04	108.2246	106.3846	110.0645	76.9000	84.7000
2014.05	110.9832	109.2094	112.7570	72.3000	75.2000
2014.06	89.9918	88.5299	91.4536	67.2000	71.0000
2014.07	98.8831	97.2708	100.4954	72.5000	72.5000
2014.08	99.4356	97.9418	100.9293	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	114.1498	112.3141	115.9855	83.2000	87.6000
2014.10	108.8806	107.0544	110.7068	59.5000	60.6000
2014.11	115.4606	113.2773	117.6439	65.8000	71.1000
2014.12	105.5946	103.3139	107.8752	75.8000	78.0000
2015.01	62.8919	61.6848	64.0990	65.9000	67.0000
2015.02	54.4254	53.1616	55.6892	42.4000	44.8000
2015.03	59.4835	58.3959	60.5711	38.0000	38.4000
2015.04	65.3714	64.2144	66.5284	49.0000	54.4000
2015.05	66.6555	65.5853	67.7257	56.3000	58.8000
2015.06	54.4967	53.5726	55.4208	50.2000	68.3000
2015.07	58.6158	57.6176	59.6141	47.9000	65.8000
2015.08	60.3689	59.3832	61.3546	39.5000	57.2000
2015.09	69.0757	67.9516	70.1998	49.2000	72.1000
2015.10	65.8336	64.7090	66.9583	39.3000	48.3000
2015.11	70.3353	69.3815	71.2890	39.6000	55.9000
2015.12	63.8210	62.5292	65.1129	36.4000	44.8000
2016.01	39.8688	39.1620	40.5756	33.7000	43.3000
2016.02	33.7706	33.1019	34.4393	38.3000	46.8000
2016.03	36.7729	36.1002	37.4456	30.5000	38.9000
2016.04	39.7401	39.0353	40.4449	26.6000	30.9000
2016.05	41.5746	40.8627	42.2865	33.7000	48.4000
2016.06	33.6332	33.0994	34.1670	13.1000	19.5000
2016.07	37.0027	36.4308	37.5745	21.2000	27.5000
2016.08	37.6938	37.0720	38.3156	33.0000	47.9000

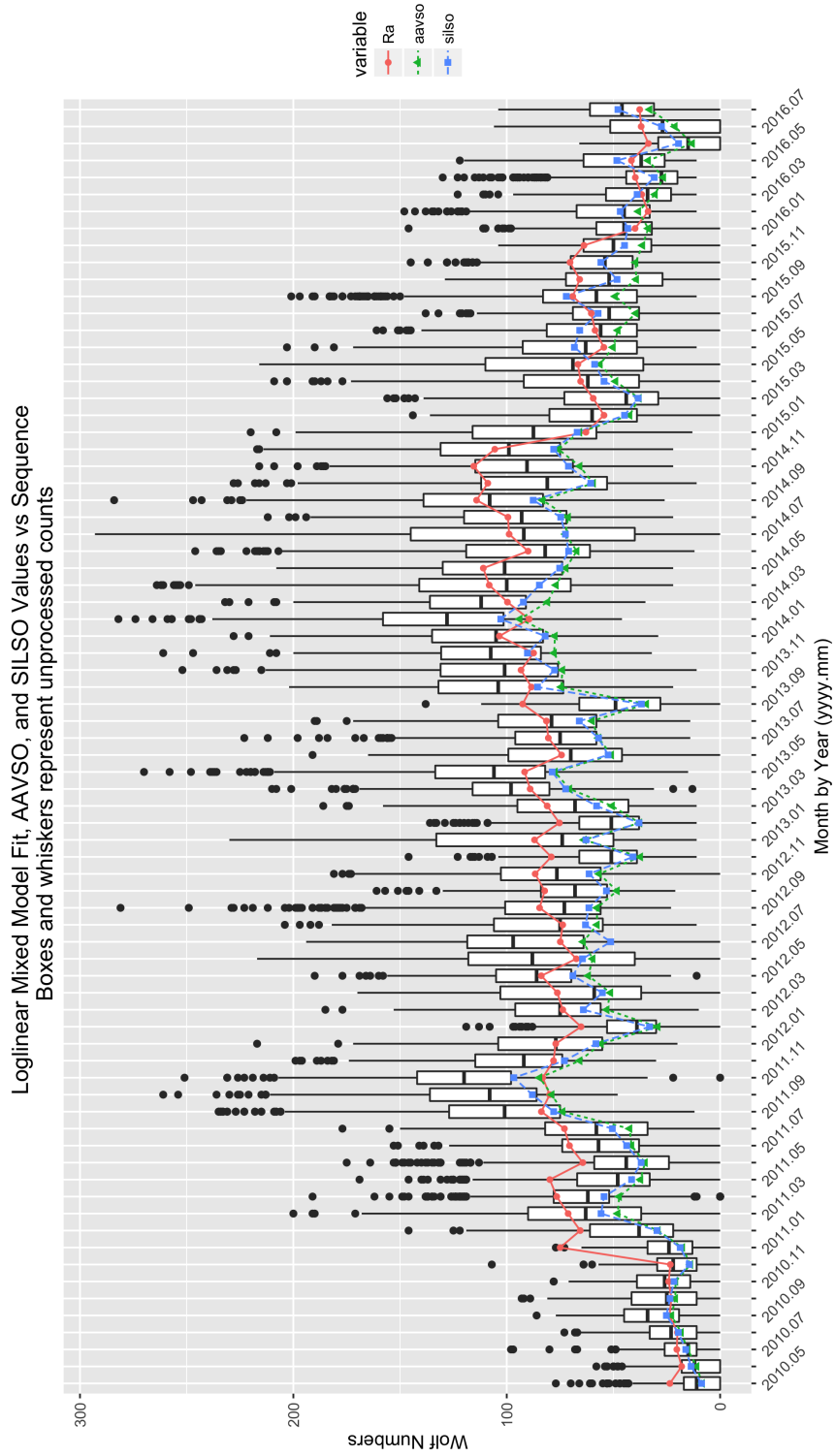


Figure 2: GLMM fitted data for R_a . AAVSO data: <https://www.aavso.org/category/tags/solar-bulletin>. SILSO data: WDC-SILSO, Royal Observatory of Belgium, Brussels

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: 201608 Parameter Estimates

	Estimate	Std. Error	t-value	$\Pr(> t)$
(Intercept)	3.2017	0.0444	72.1683	0.0000
seeF	-0.1885	0.0073	-25.9931	0.0000
seeG	-0.1014	0.0063	-16.0808	0.0000
seeP	-0.2946	0.0107	-27.6555	0.0000
r1000B	-0.0592	0.0828	-0.7140	0.4752
r1500C	0.0310	0.1266	0.2449	0.8065
r2000D	0.0779	0.1544	0.5046	0.6138
r2500E	-0.0007	0.1049	-0.0068	0.9946
r3000F	0.0697	0.1021	0.6830	0.4946
r3500G	0.1213	0.1528	0.7935	0.4275
r5000H	-0.1100	0.2114	-0.5203	0.6028
silsoy	0.1195	0.0736	1.6233	0.1045
year2011	1.2104	0.0153	79.0083	0.0000
year2012	1.2276	0.0153	80.3876	0.0000
year2013	1.3249	0.0152	87.0061	0.0000
year2014	1.5141	0.0151	100.0915	0.0000
year2015	1.0135	0.0155	65.3092	0.0000
year2016	0.5367	0.0172	31.1271	0.0000
mon2	-0.1536	0.0120	-12.7533	0.0000
mon3	-0.0645	0.0111	-5.8219	0.0000
mon4	0.0210	0.0111	1.8870	0.0592
mon5	0.0489	0.0105	4.6412	0.0000
mon6	-0.1699	0.0111	-15.2391	0.0000
mon7	-0.0811	0.0107	-7.5829	0.0000
mon8	-0.0646	0.0105	-6.1337	0.0000
mon9	0.0745	0.0103	7.2346	0.0000
mon10	0.0330	0.0109	3.0318	0.0024
mon11	0.0957	0.0111	8.6286	0.0000
mon12	0.0169	0.0118	1.4361	0.1510

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.

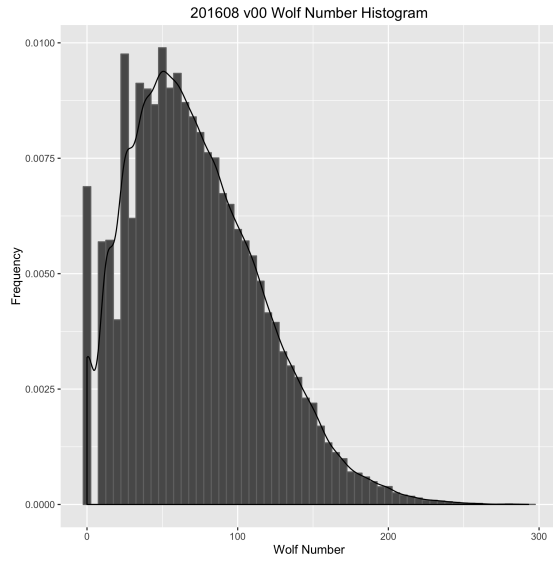
5 Supporting Information

Table 4: 201608 Summary of Sunspot Numbers

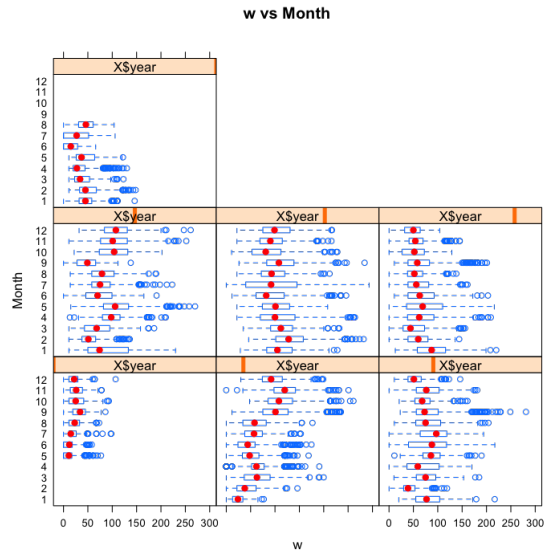
obs	jd	year	mon	day
ARAG : 2260	Min. :1721096	Min. :2010	Min. : 1.000	Min. : 1.00
CHAG : 2079	1st Qu.:2455971	1st Qu.:2012	1st Qu.: 4.000	1st Qu.: 8.00
BRAB : 2041	Median :2456507	Median :2013	Median : 7.000	Median :16.00
BROB : 1839	Mean :2456180	Mean :2013	Mean : 6.608	Mean :15.74
DUBF : 1728	3rd Qu.:2457072	3rd Qu.:2015	3rd Qu.: 9.000	3rd Qu.:23.00
HOWR : 1699	Max. :2457632	Max. :2016	Max. :12.000	Max. :31.00
(Other):41400				

Table 5: Summary of Sunspot Numbers

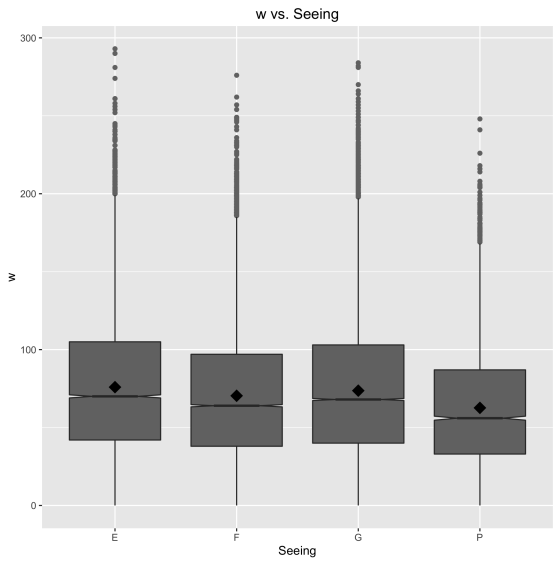
see	g	s	w	r	silso
E: 9768	Min. : 0.000	Min. : 0.00	Min. : 0.00	0000A :22606	n:35735
F:16252	1st Qu.: 3.000	1st Qu.: 10.00	1st Qu.: 39.00	3000F : 8675	y:17311
G:22693	Median : 4.000	Median : 22.00	Median : 66.00	2500E : 6919	
P: 4333	Mean : 4.443	Mean : 27.74	Mean : 72.17	3500G : 4120	
	3rd Qu.: 6.000	3rd Qu.: 40.00	3rd Qu.:100.00	1000B : 3865	
	Max. :18.000	Max. :204.00	Max. :293.00	1500C : 3006	
				(Other): 3855	



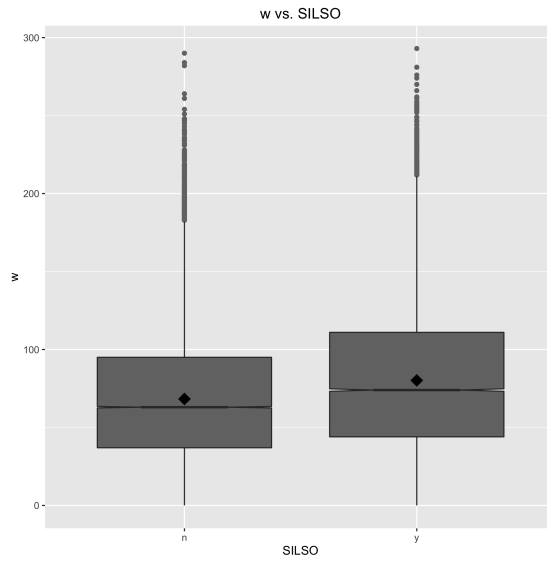
(a) Observed sunspot count histogram.



(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:

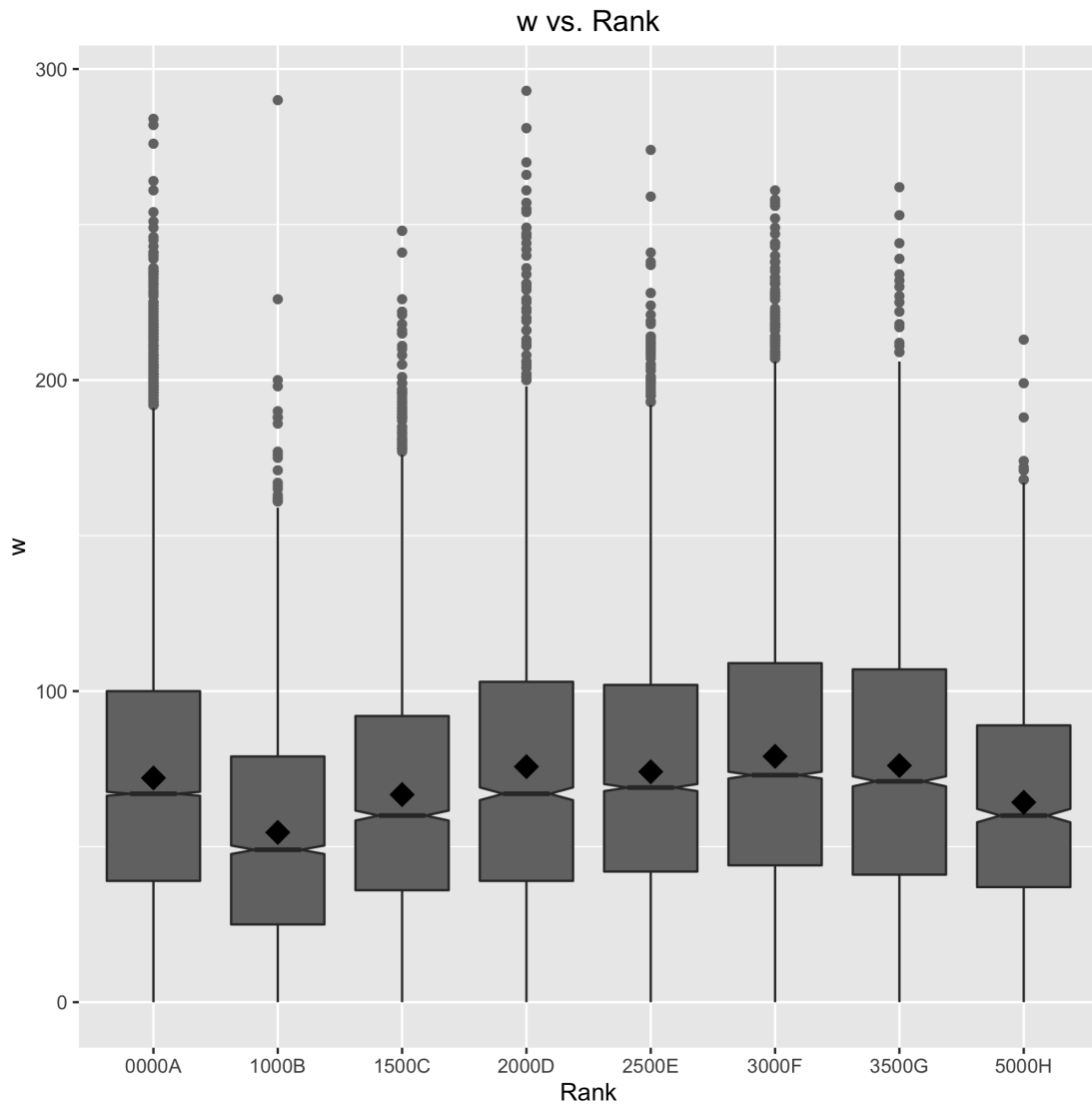


Figure 4: Box plot of sunspot count by rank.