

Monthly Report (00)

2016.06 Data Set

Friday 22nd July, 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.

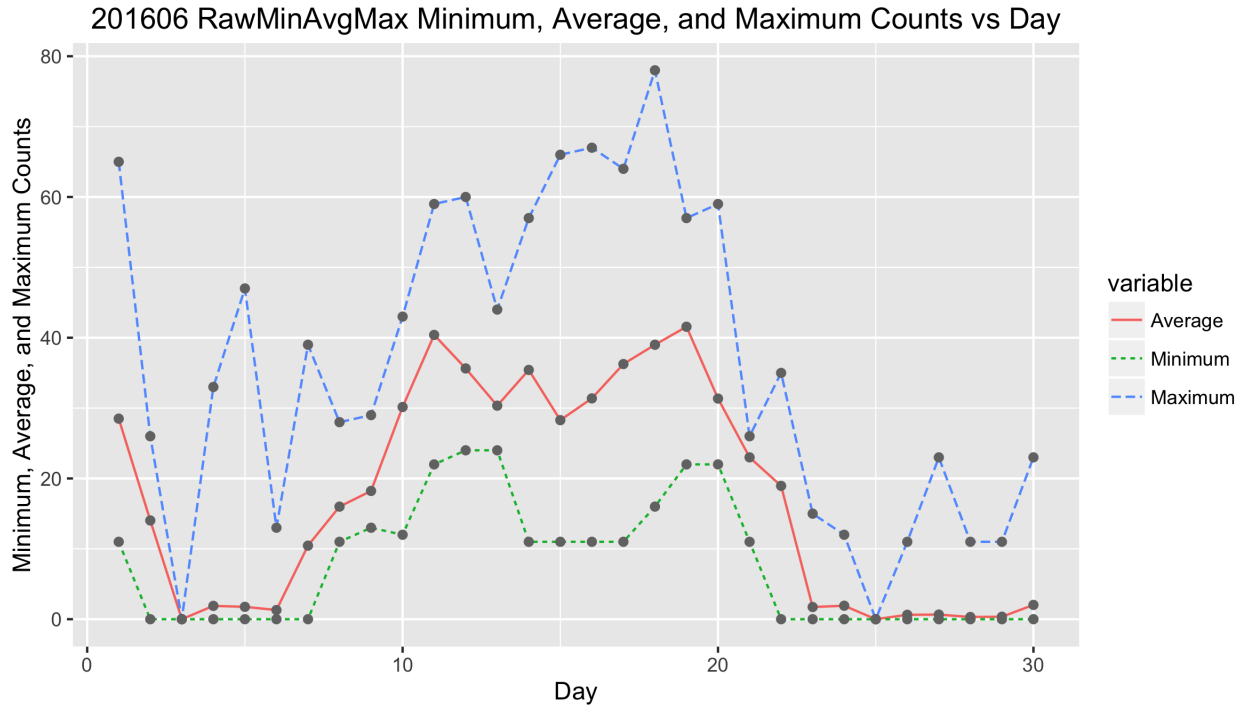


Figure 1: Raw average sunspot count by day of the month.

Table 1: 201606 Daily Raw Counts

Day	Submissions	Minimum	Average	Maximum
1.0000	38.0000	11.0000	28.5000	65.0000
2.0000	37.0000	0.0000	14.0323	26.0000
3.0000	32.0000	0.0000	0.0000	0.0000
4.0000	32.0000	0.0000	1.8966	33.0000
5.0000	35.0000	0.0000	1.7576	47.0000
6.0000	37.0000	0.0000	1.3056	13.0000
7.0000	35.0000	0.0000	10.4545	39.0000
8.0000	41.0000	11.0000	16.0000	28.0000
9.0000	36.0000	13.0000	18.2286	29.0000
10.0000	40.0000	12.0000	30.1538	43.0000
11.0000	35.0000	22.0000	40.4000	59.0000
12.0000	34.0000	24.0000	35.6333	60.0000
13.0000	33.0000	24.0000	30.3438	44.0000
14.0000	38.0000	11.0000	35.4211	57.0000
15.0000	38.0000	11.0000	28.2857	66.0000
16.0000	30.0000	11.0000	31.3793	67.0000
17.0000	35.0000	11.0000	36.2581	64.0000
18.0000	38.0000	16.0000	39.0000	78.0000
19.0000	39.0000	22.0000	41.5625	57.0000
20.0000	38.0000	22.0000	31.3514	59.0000
21.0000	40.0000	11.0000	23.0000	26.0000
22.0000	37.0000	0.0000	18.9429	35.0000
23.0000	43.0000	0.0000	1.7297	15.0000
24.0000	39.0000	0.0000	1.9167	12.0000
25.0000	34.0000	0.0000	0.0000	0.0000
26.0000	40.0000	0.0000	0.6286	11.0000
27.0000	39.0000	0.0000	0.6571	23.0000
28.0000	37.0000	0.0000	0.3143	11.0000
29.0000	33.0000	0.0000	0.3548	11.0000
30.0000	43.0000	0.0000	2.0286	23.0000

3 Error Tables

Data are for the month of June 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	23.5338	22.9952	24.0723	8.4000	8.7000
2010.06	18.0178	17.5356	18.5000	11.0000	13.6000
2010.07	20.6539	20.1979	21.1099	15.2000	16.1000
2010.08	19.7055	19.2249	20.1862	18.3000	19.6000
2010.09	23.7246	23.2156	24.2336	22.8000	25.2000
2010.10	22.7861	22.2942	23.2779	21.0000	23.5000
2010.11	24.1040	23.5582	24.6498	20.9000	21.6000
2010.12	23.3198	22.6485	23.9911	13.9000	14.5000
2011.01	74.9587	73.2748	76.6426	17.7000	18.7000
2011.02	65.5416	64.0599	67.0233	29.1000	29.6000
2011.03	71.2250	69.7551	72.6950	48.0000	55.8000
2011.04	76.5375	74.8982	78.1769	47.3000	54.4000
2011.05	79.6971	78.1017	81.2926	37.3000	41.5000
2011.06	64.3119	62.9600	65.6638	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	71.8535	70.2752	73.4318	41.5000	43.8000
2011.08	71.8598	70.4500	73.2695	42.4000	50.5000
2011.09	83.7995	82.7012	84.8978	73.8000	78.0000
2011.10	79.9463	78.5773	81.3152	78.9000	88.0000
2011.11	83.1696	81.3918	84.9474	84.6000	96.7000
2011.12	78.1237	76.4162	79.8311	65.8000	73.0000
2012.01	77.0292	75.5011	78.5574	55.8000	58.2000
2012.02	65.1571	63.7763	66.5380	29.2000	33.1000
2012.03	73.7334	72.4176	75.0492	53.1000	64.1000
2012.04	76.2702	74.0346	78.5058	51.4000	55.2000
2012.05	83.7716	82.3084	85.2349	61.8000	69.0000
2012.06	67.3883	66.1900	68.5866	59.7000	64.5000
2012.07	76.2709	74.9889	77.5528	64.2000	51.3000
2012.08	72.6486	71.4341	73.8631	57.7000	63.1000
2012.09	84.6135	83.1679	86.0591	57.7000	61.5000
2012.10	82.2357	80.6829	83.7885	48.3000	53.3000
2012.11	86.6630	84.9261	88.3999	56.7000	61.4000
2012.12	79.1164	77.4451	80.7877	37.4000	40.8000
2013.01	86.9641	85.3284	88.5998	63.8000	62.9000
2013.02	75.2803	73.8135	76.7470	37.8000	38.0000
2013.03	80.9960	79.4716	82.5204	50.6000	57.9000
2013.04	89.0050	87.5031	90.5070	70.6000	72.4000
2013.05	91.5973	90.0138	93.1809	77.4000	78.7000
2013.06	74.2924	72.9667	75.6181	51.0000	52.5000
2013.07	81.9693	80.6895	83.2491	57.0000	57.0000
2013.08	80.1725	78.9183	81.4267	60.0000	66.0000
2013.09	92.5385	90.9369	94.1400	34.6000	36.9000
2013.10	88.4287	86.8557	90.0017	74.5000	85.6000
2013.11	93.3189	91.3711	95.2666	73.9000	77.6000
2013.12	87.4448	85.6840	89.2055	77.8000	90.3000
2014.01	103.4244	101.2492	105.5996	77.4000	82.0000
2014.02	89.4966	87.7940	91.1993	93.9000	102.8000
2014.03	99.6455	97.9633	101.3277	80.9000	92.2000
2014.04	108.1626	106.3216	110.0036	76.9000	84.7000
2014.05	110.9448	109.1718	112.7178	72.3000	75.2000
2014.06	89.9679	88.5041	91.4316	67.2000	71.0000
2014.07	100.6891	99.0491	102.3292	72.5000	72.5000
2014.08	97.9872	96.5085	99.4660	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	114.1323	112.3002	115.9644	83.2000	87.6000
2014.10	108.8646	107.0394	110.6898	59.5000	60.6000
2014.11	115.4388	113.2523	117.6254	65.8000	71.1000
2014.12	105.6122	103.3333	107.8912	75.8000	78.0000
2015.01	62.8432	61.6317	64.0546	65.9000	67.0000
2015.02	54.4001	53.1378	55.6624	42.4000	44.8000
2015.03	59.4631	58.3753	60.5510	38.0000	38.4000
2015.04	65.3640	64.2083	66.5197	49.0000	54.4000
2015.05	66.7305	65.6532	67.8079	56.3000	58.8000
2015.06	54.4917	53.5674	55.4161	50.2000	68.3000
2015.07	59.6653	58.6491	60.6815	47.9000	65.8000
2015.08	59.4146	58.4423	60.3870	39.5000	57.2000
2015.09	69.0521	67.9278	70.1764	49.2000	72.1000
2015.10	65.8271	64.7017	66.9524	39.3000	48.3000
2015.11	70.3386	69.3842	71.2930	39.6000	55.9000
2015.12	63.7610	62.4635	65.0585	36.4000	44.8000
2016.01	40.0181	39.3083	40.7279	33.7000	43.3000
2016.02	33.8926	33.2204	34.5648	38.3000	46.8000
2016.03	36.9132	36.2379	37.5884	30.5000	38.9000
2016.04	39.8944	39.1855	40.6033	26.6000	30.9000
2016.05	41.7157	40.9998	42.4317	33.7000	48.4000
2016.06	33.7608	33.2230	34.2986	13.1000	19.5000

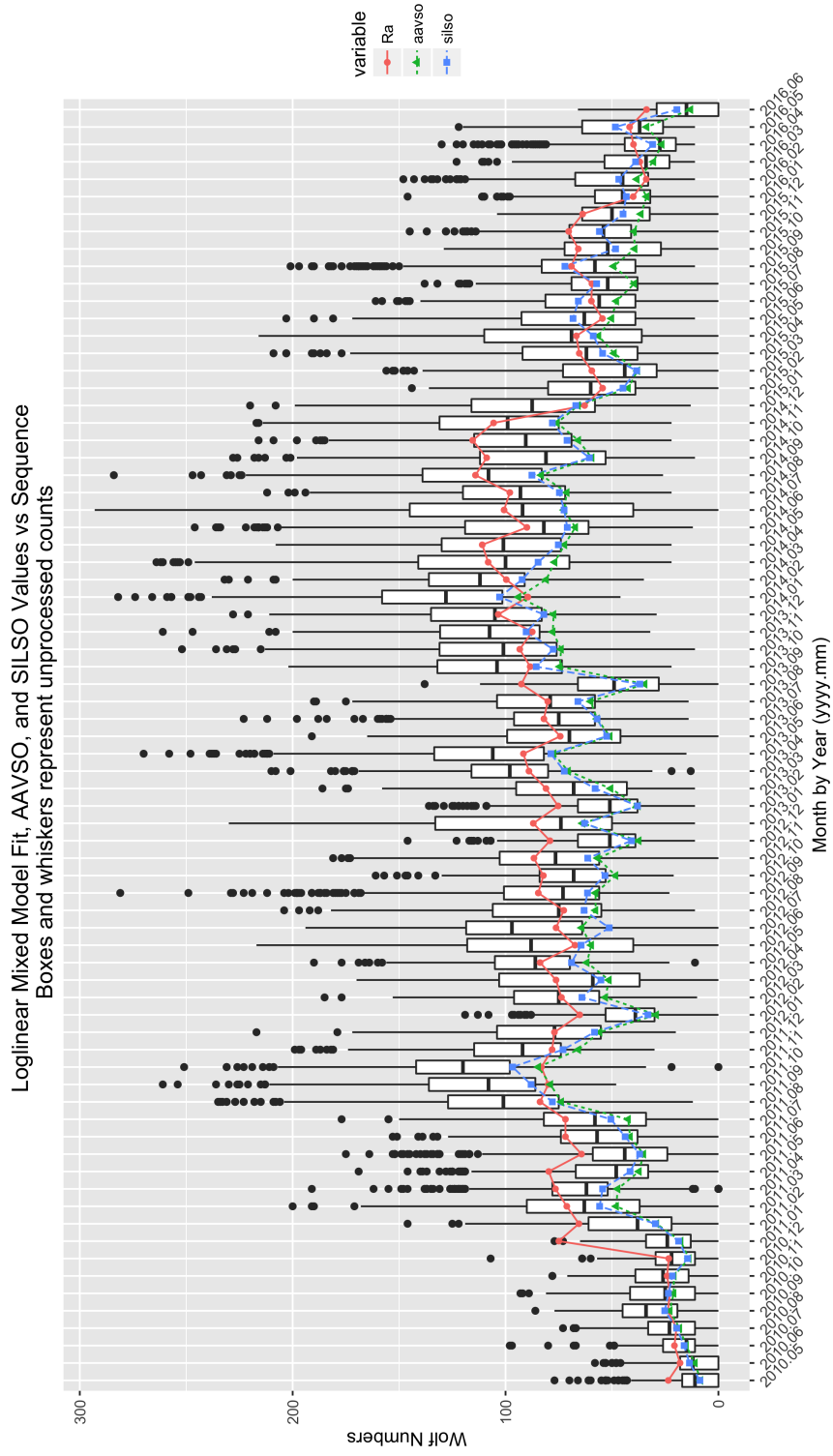


Figure 2: GLMM fitted data for R_a . AAVSO data: <https://www.aavso.org/category/tags/solar-bulletin>. SILSO data: WDC-SILSO, Royal Observatory of Belgium, Brussels

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: 201606 Parameter Estimates

	Estimate	Std. Error	t-value	$\Pr(> t)$
(Intercept)	3.2013	0.0446	71.8120	0.0000
seeF	-0.1849	0.0073	-25.4213	0.0000
seeG	-0.0970	0.0063	-15.3248	0.0000
seeP	-0.2922	0.0107	-27.4003	0.0000
r1000B	-0.0653	0.0833	-0.7838	0.4332
r1500C	0.0277	0.1274	0.2171	0.8282
r2000D	0.0749	0.1554	0.4821	0.6297
r2500E	-0.0019	0.1055	-0.0180	0.9857
r3000F	0.0663	0.1028	0.6452	0.5188
r3500G	0.1194	0.1538	0.7764	0.4375
r5000H	-0.1144	0.2127	-0.5381	0.5905
silsoy	0.1174	0.0741	1.5843	0.1131
year2011	1.2110	0.0152	79.4545	0.0000
year2012	1.2279	0.0152	80.8236	0.0000
year2013	1.3253	0.0151	87.4800	0.0000
year2014	1.5144	0.0150	100.6237	0.0000
year2015	1.0138	0.0154	65.6668	0.0000
year2016	0.5412	0.0182	29.7240	0.0000
mon2	-0.1536	0.0120	-12.8238	0.0000
mon3	-0.0644	0.0110	-5.8500	0.0000
mon4	0.0214	0.0111	1.9286	0.0538
mon5	0.0491	0.0105	4.6854	0.0000
mon6	-0.1693	0.0111	-15.2696	0.0000
mon7	-0.0624	0.0108	-5.7582	0.0000
mon8	-0.0793	0.0107	-7.3913	0.0000
mon9	0.0750	0.0103	7.3142	0.0000
mon10	0.0336	0.0108	3.0947	0.0020
mon11	0.0960	0.0110	8.6965	0.0000
mon12	0.0172	0.0117	1.4663	0.1426

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.

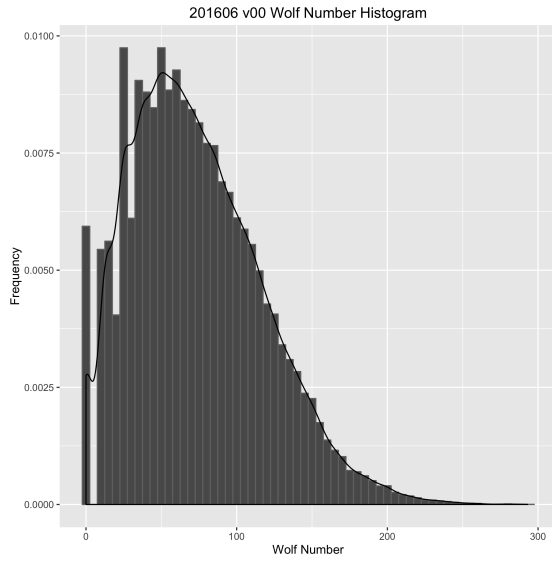
5 Supporting Information

Table 4: 201606 Summary of Sunspot Numbers

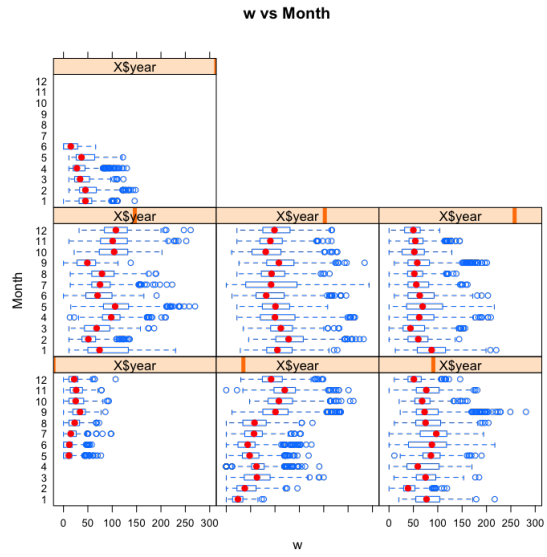
obs	jd	year	mon	day
ARAG : 2198	Min. :1721096	Min. :2010	Min. : 1.00	Min. : 1.00
CHAG : 2018	1st Qu.:2455954	1st Qu.:2012	1st Qu.: 4.00	1st Qu.: 8.00
BRAB : 1981	Median :2456481	Median :2013	Median : 7.00	Median :16.00
BROB : 1777	Mean :2456137	Mean :2013	Mean : 6.58	Mean :15.73
DUBF : 1670	3rd Qu.:2457002	3rd Qu.:2014	3rd Qu.: 9.00	3rd Qu.:23.00
HOWR : 1645	Max. :2457570	Max. :2016	Max. :12.00	Max. :31.00
(Other):40174				

Table 5: Summary of Sunspot Numbers

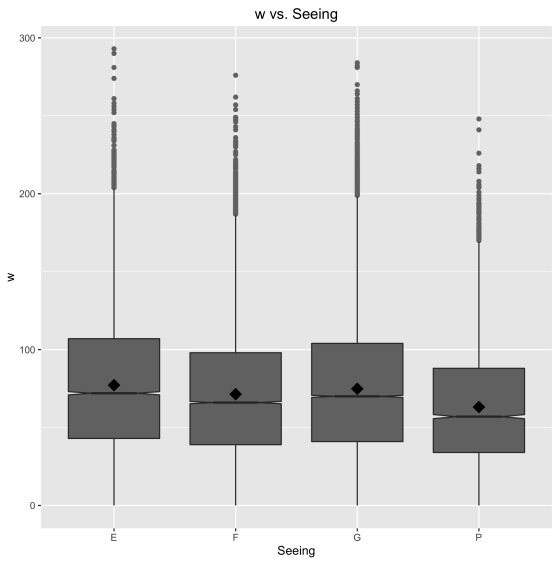
see	g	s	w	r	silso
E: 9404	Min. : 0.000	Min. : 0.0	Min. : 0.00	0000A :22008	n:34667
F:15817	1st Qu.: 3.000	1st Qu.: 10.0	1st Qu.: 40.00	3000F : 8388	y:16796
G:21987	Median : 4.000	Median : 22.0	Median : 68.00	2500E : 6681	
P: 4255	Mean : 4.506	Mean : 28.2	Mean : 73.25	3500G : 3999	
	3rd Qu.: 6.000	3rd Qu.: 40.0	3rd Qu.:101.00	1000B : 3713	
	Max. :18.000	Max. :204.0	Max. :293.00	1500C : 2962	
				(Other): 3712	



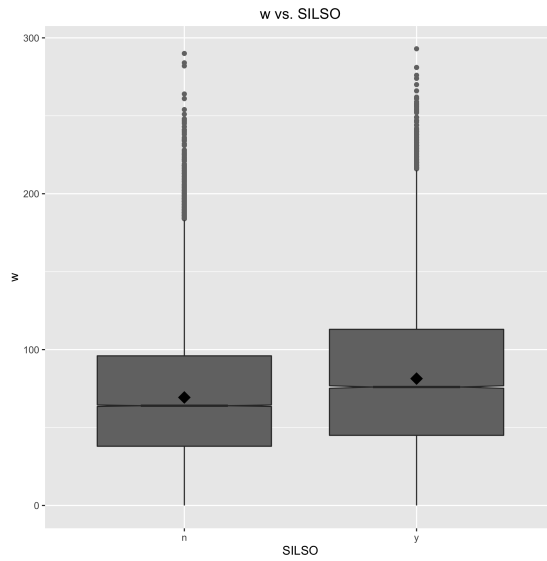
(a) Observed sunspot count histogram.



(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:

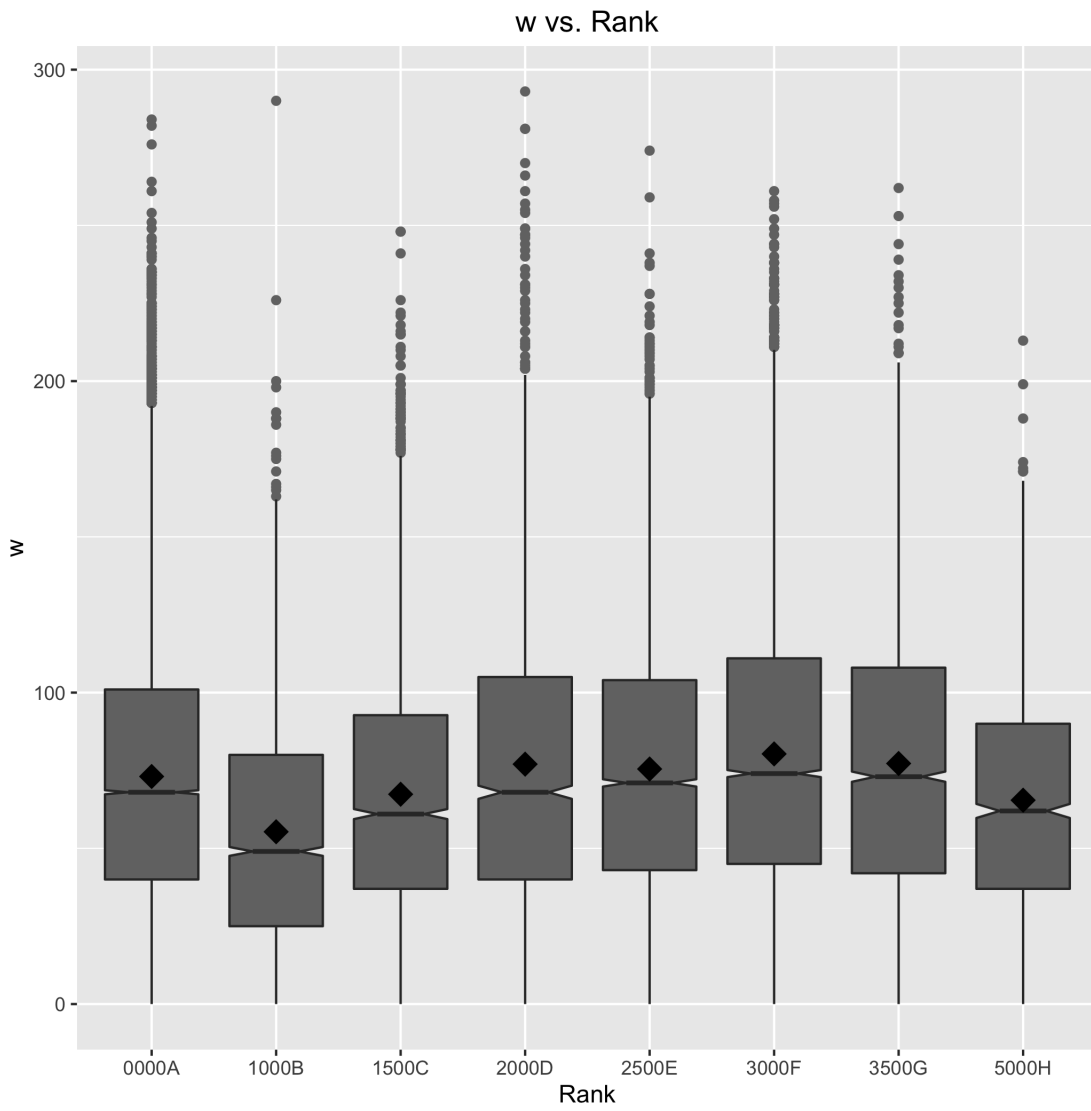


Figure 4: Box plot of sunspot count by rank.