

Monthly Report (00)

2016.07 Data Set

Saturday 13th August, 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.

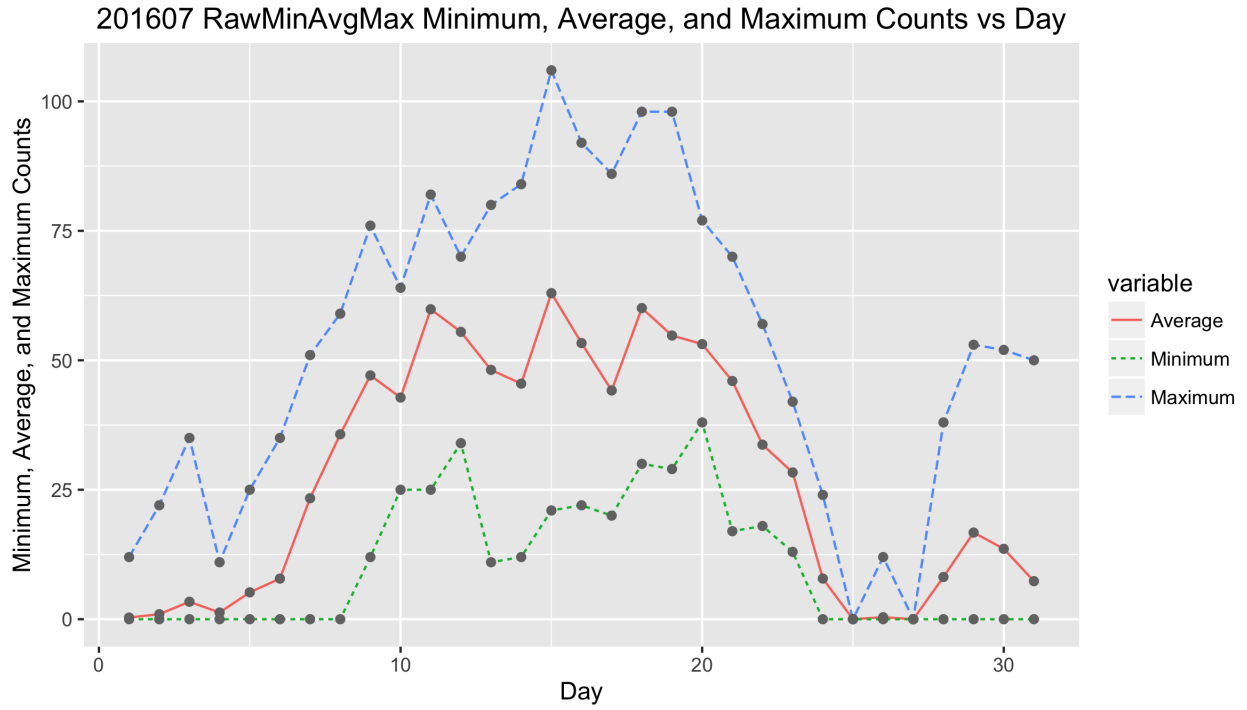


Figure 1: Raw average sunspot count by day of the month.

Table 1: 201607 Daily Raw Counts

Day	Submissions	Minimum	Average	Maximum
1.0000	39.0000	0.0000	0.3243	12.0000
2.0000	37.0000	0.0000	0.9706	22.0000
3.0000	40.0000	0.0000	3.3714	35.0000
4.0000	37.0000	0.0000	1.2941	11.0000
5.0000	36.0000	0.0000	5.1875	25.0000
6.0000	33.0000	0.0000	7.8387	35.0000
7.0000	36.0000	0.0000	23.3636	51.0000
8.0000	34.0000	0.0000	35.7188	59.0000
9.0000	39.0000	12.0000	47.0833	76.0000
10.0000	40.0000	25.0000	42.8056	64.0000
11.0000	36.0000	25.0000	59.8571	82.0000
12.0000	38.0000	34.0000	55.4857	70.0000
13.0000	38.0000	11.0000	48.1429	80.0000
14.0000	36.0000	12.0000	45.4848	84.0000
15.0000	32.0000	21.0000	62.9655	106.0000
16.0000	39.0000	22.0000	53.3235	92.0000
17.0000	37.0000	20.0000	44.1714	86.0000
18.0000	41.0000	30.0000	60.0789	98.0000
19.0000	37.0000	29.0000	54.7812	98.0000
20.0000	36.0000	38.0000	53.1250	77.0000
21.0000	35.0000	17.0000	46.0000	70.0000
22.0000	40.0000	18.0000	33.7027	57.0000
23.0000	38.0000	13.0000	28.3429	42.0000
24.0000	42.0000	0.0000	7.8462	24.0000
25.0000	36.0000	0.0000	0.0000	0.0000
26.0000	37.0000	0.0000	0.3871	12.0000
27.0000	34.0000	0.0000	0.0000	0.0000
28.0000	35.0000	0.0000	8.1515	38.0000
29.0000	32.0000	0.0000	16.7333	53.0000
30.0000	35.0000	0.0000	13.5882	52.0000
31.0000	36.0000	0.0000	7.3438	50.0000

3 Error Tables

Data are for the month of July 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	23.6368	23.0953	24.1783	8.4000	8.7000
2010.06	18.1033	17.6187	18.5879	11.0000	13.6000
2010.07	20.3695	19.9192	20.8199	15.2000	16.1000
2010.08	19.7355	19.2539	20.2172	18.3000	19.6000
2010.09	23.7603	23.2501	24.2706	22.8000	25.2000
2010.10	22.8175	22.3244	23.3106	21.0000	23.5000
2010.11	24.1386	23.5919	24.6853	20.9000	21.6000
2010.12	23.3593	22.6858	24.0328	13.9000	14.5000
2011.01	75.1793	73.4893	76.8693	17.7000	18.7000
2011.02	65.7393	64.2521	67.2265	29.1000	29.6000
2011.03	71.4275	69.9520	72.9031	48.0000	55.8000
2011.04	76.8107	75.1635	78.4579	47.3000	54.4000
2011.05	79.9482	78.3457	81.5508	37.3000	41.5000
2011.06	64.5187	63.1594	65.8780	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	70.7674	69.2106	72.3242	41.5000	43.8000
2011.08	71.8568	70.4459	73.2677	42.4000	50.5000
2011.09	83.7786	82.6785	84.8787	73.8000	78.0000
2011.10	79.9131	78.5413	81.2849	78.9000	88.0000
2011.11	83.1278	81.3485	84.9071	84.6000	96.7000
2011.12	78.0860	76.3758	79.7962	65.8000	73.0000
2012.01	77.2885	75.7541	78.8229	55.8000	58.2000
2012.02	65.3765	63.9893	66.7637	29.2000	33.1000
2012.03	73.9592	72.6375	75.2810	53.1000	64.1000
2012.04	76.5922	74.3345	78.8498	51.4000	55.2000
2012.05	84.0494	82.5794	85.5194	61.8000	69.0000
2012.06	67.6186	66.4150	68.8222	59.7000	64.5000
2012.07	75.1368	73.8712	76.4024	64.2000	51.3000
2012.08	72.6697	71.4532	73.8862	57.7000	63.1000
2012.09	84.6215	83.1742	86.0689	57.7000	61.5000
2012.10	82.2463	80.6918	83.8008	48.3000	53.3000
2012.11	86.6908	84.9523	88.4294	56.7000	61.4000
2012.12	79.1272	77.4543	80.8001	37.4000	40.8000
2013.01	87.2292	85.5883	88.8701	63.8000	62.9000
2013.02	75.5156	74.0427	76.9885	37.8000	38.0000
2013.03	81.2439	79.7125	82.7752	50.6000	57.9000
2013.04	89.3323	87.8236	90.8411	70.6000	72.4000
2013.05	91.8763	90.2866	93.4659	77.4000	78.7000
2013.06	74.5315	73.2003	75.8627	51.0000	52.5000
2013.07	80.7296	79.4691	81.9900	57.0000	57.0000
2013.08	80.1817	78.9267	81.4368	60.0000	66.0000
2013.09	92.5342	90.9321	94.1363	34.6000	36.9000
2013.10	88.4346	86.8608	90.0084	74.5000	85.6000
2013.11	93.3061	91.3581	95.2542	73.9000	77.6000
2013.12	87.4305	85.6696	89.1913	77.8000	90.3000
2014.01	103.7278	101.5447	105.9108	77.4000	82.0000
2014.02	89.7736	88.0632	91.4840	93.9000	102.8000
2014.03	99.9474	98.2588	101.6361	80.9000	92.2000
2014.04	108.5532	106.7033	110.4032	76.9000	84.7000
2014.05	111.2712	109.4909	113.0515	72.3000	75.2000
2014.06	90.2546	88.7853	91.7238	67.2000	71.0000
2014.07	99.1615	97.5435	100.7795	72.5000	72.5000
2014.08	97.9277	96.4530	99.4023	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	114.1413	112.3045	115.9780	83.2000	87.6000
2014.10	108.8596	107.0328	110.6865	59.5000	60.6000
2014.11	115.4424	113.2563	117.6286	65.8000	71.1000
2014.12	105.6115	103.3285	107.8945	75.8000	78.0000
2015.01	63.0445	61.8287	64.2602	65.9000	67.0000
2015.02	54.5721	53.3038	55.8405	42.4000	44.8000
2015.03	59.6317	58.5393	60.7241	38.0000	38.4000
2015.04	65.5909	64.4288	66.7530	49.0000	54.4000
2015.05	66.8466	65.7708	67.9224	56.3000	58.8000
2015.06	54.6603	53.7318	55.5889	50.2000	68.3000
2015.07	58.7698	57.7672	59.7724	47.9000	65.8000
2015.08	59.4288	58.4548	60.4028	39.5000	57.2000
2015.09	69.0607	67.9343	70.1870	49.2000	72.1000
2015.10	65.8255	64.6989	66.9522	39.3000	48.3000
2015.11	70.3341	69.3787	71.2896	39.6000	55.9000
2015.12	63.7831	62.4854	65.0808	36.4000	44.8000
2016.01	38.6412	37.9548	39.3276	33.7000	43.3000
2016.02	32.7320	32.0823	33.3818	38.3000	46.8000
2016.03	35.6354	34.9825	36.2883	30.5000	38.9000
2016.04	38.5352	37.8496	39.2209	26.6000	30.9000
2016.05	40.2777	39.5857	40.9697	33.7000	48.4000
2016.06	32.5941	32.0745	33.1136	13.1000	19.5000
2016.07	35.8541	35.2972	36.4109	21.2000	27.5000

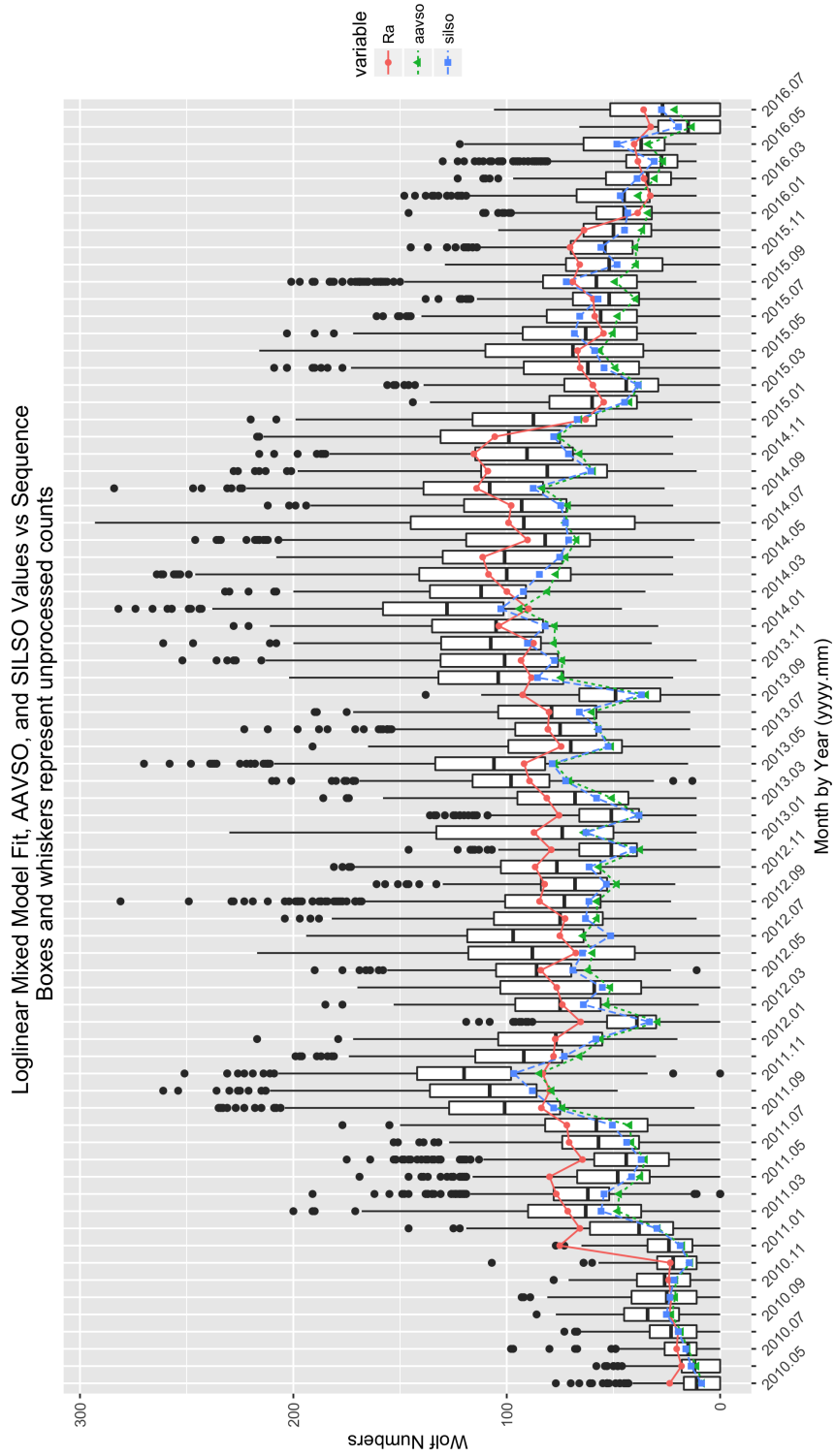


Figure 2: GLMM fitted data for R_a . AAVSO data: <https://www.aavso.org/category/tags/solar-bulletin>. SILSO data: WDC-SILSO, Royal Observatory of Belgium, Brussels

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: 201607 Parameter Estimates

	Estimate	Std. Error	t-value	$\Pr(> t)$
(Intercept)	3.2051	0.0444	72.1915	0.0000
seeF	-0.1878	0.0073	-25.7736	0.0000
seeG	-0.1000	0.0063	-15.7747	0.0000
seeP	-0.2939	0.0107	-27.4790	0.0000
r1000B	-0.0614	0.0829	-0.7405	0.4590
r1500C	0.0310	0.1267	0.2449	0.8065
r2000D	0.0769	0.1545	0.4974	0.6189
r2500E	-0.0014	0.1049	-0.0136	0.9892
r3000F	0.0690	0.1022	0.6751	0.4996
r3500G	0.1214	0.1529	0.7942	0.4271
r5000H	-0.1113	0.2115	-0.5264	0.5986
silsoy	0.1199	0.0737	1.6284	0.1034
year2011	1.2093	0.0153	78.9303	0.0000
year2012	1.2268	0.0153	80.3241	0.0000
year2013	1.3239	0.0152	86.9312	0.0000
year2014	1.5130	0.0151	100.0049	0.0000
year2015	1.0123	0.0155	65.2276	0.0000
year2016	0.5015	0.0178	28.2097	0.0000
mon2	-0.1535	0.0120	-12.7497	0.0000
mon3	-0.0646	0.0111	-5.8367	0.0000
mon4	0.0217	0.0111	1.9449	0.0518
mon5	0.0488	0.0105	4.6370	0.0000
mon6	-0.1695	0.0111	-15.2047	0.0000
mon7	-0.0810	0.0107	-7.5749	0.0000
mon8	-0.0825	0.0108	-7.6591	0.0000
mon9	0.0717	0.0103	6.9601	0.0000
mon10	0.0302	0.0109	2.7755	0.0055
mon11	0.0928	0.0111	8.3667	0.0000
mon12	0.0141	0.0118	1.1935	0.2327

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.

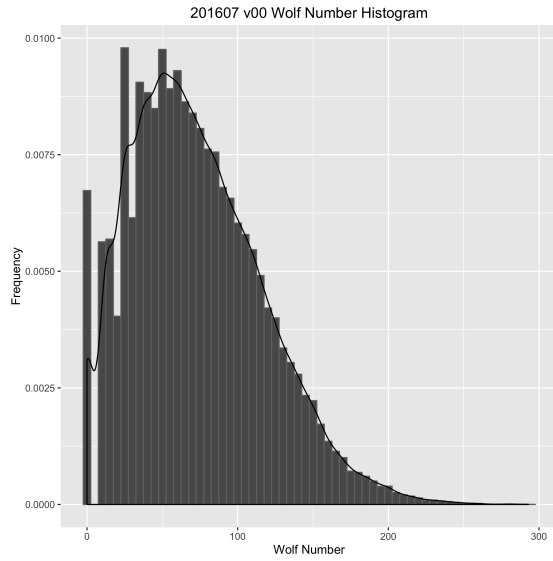
5 Supporting Information

Table 4: 201607 Summary of Sunspot Numbers

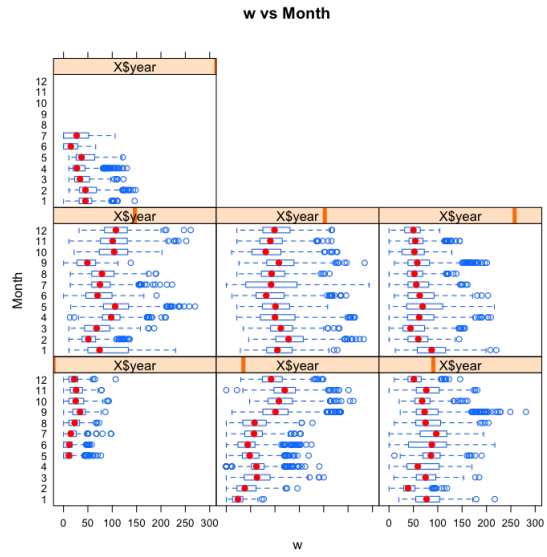
obs	jd	year	mon	day
ARAG : 2229	Min. :1721096	Min. :2010	Min. : 1.000	Min. : 1.00
CHAG : 2048	1st Qu.:2455962	1st Qu.:2012	1st Qu.: 4.000	1st Qu.: 8.00
BRAB : 2012	Median :2456494	Median :2013	Median : 7.000	Median :16.00
BROB : 1808	Mean :2456159	Mean :2013	Mean : 6.587	Mean :15.73
DUBF : 1700	3rd Qu.:2457038	3rd Qu.:2015	3rd Qu.: 9.000	3rd Qu.:23.00
HOWR : 1672	Max. :2457601	Max. :2016	Max. :12.000	Max. :31.00
(Other):40793				

Table 5: Summary of Sunspot Numbers

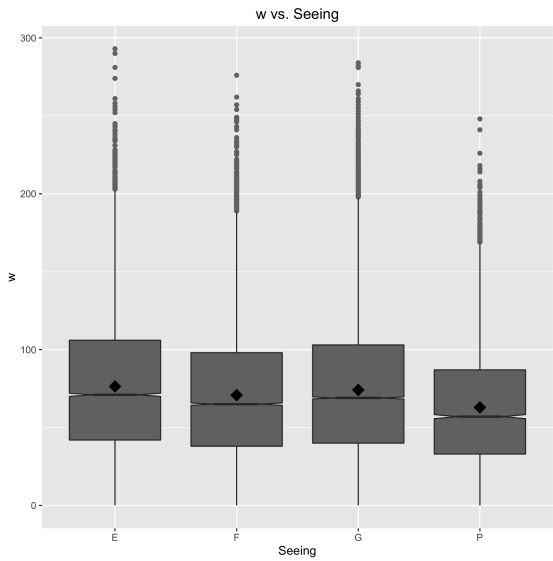
see	g	s	w	r	silso
E: 9598	Min. : 0.000	Min. : 0.00	Min. : 0.00	0000A :22313	n:35209
F:16035	1st Qu.: 3.000	1st Qu.: 10.00	1st Qu.: 39.00	3000F : 8532	y:17053
G:22345	Median : 4.000	Median : 22.00	Median : 67.00	2500E : 6800	
P: 4284	Mean : 4.465	Mean : 27.94	Mean : 72.58	3500G : 4060	
	3rd Qu.: 6.000	3rd Qu.: 40.00	3rd Qu.:101.00	1000B : 3789	
	Max. :18.000	Max. :204.00	Max. :293.00	1500C : 2987	
				(Other): 3781	



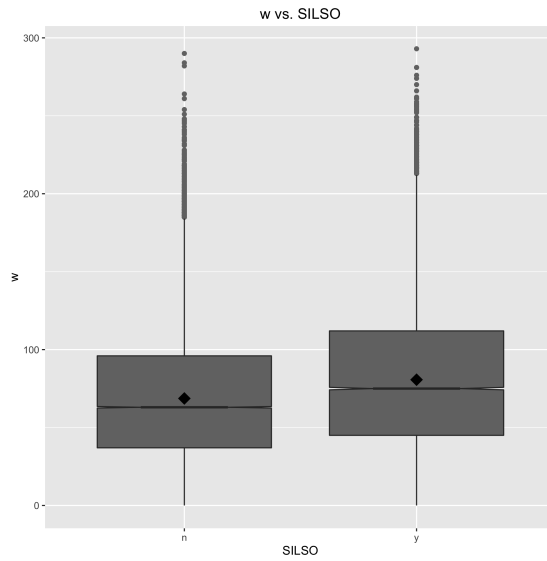
(a) Observed sunspot count histogram.



(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:

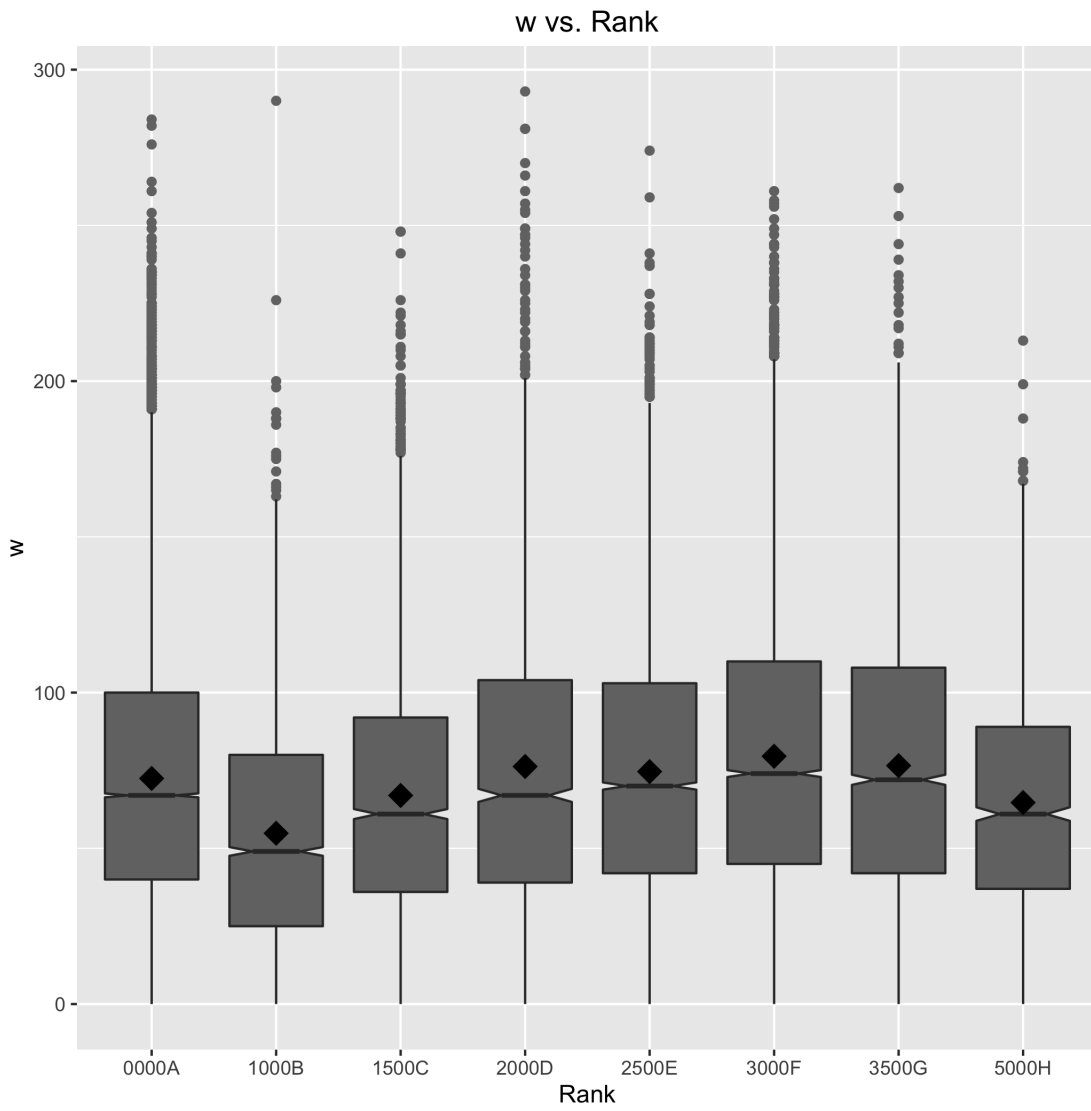


Figure 4: Box plot of sunspot count by rank.