

Monthly Report (00)

2016.05 Data Set

Tuesday 14th June, 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.

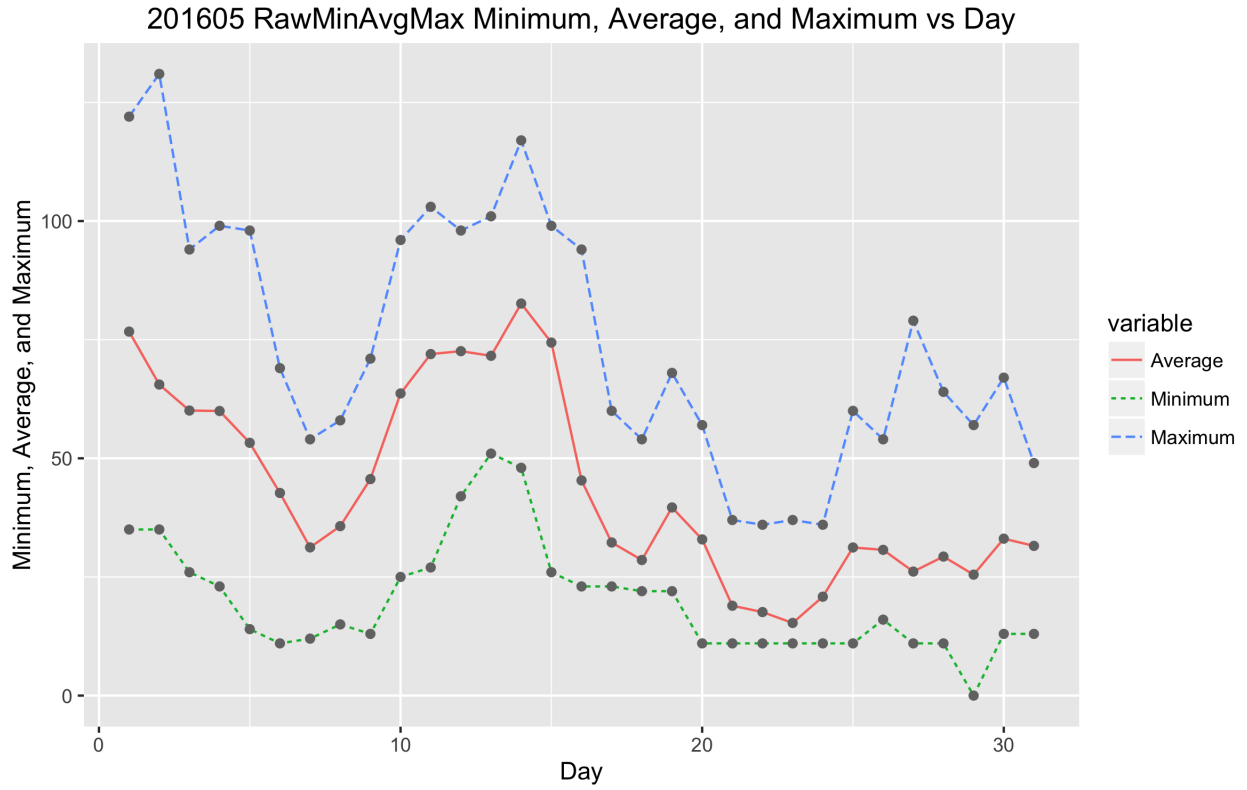


Figure 1: Raw average sunspot count by day of the month.

Table 1: 201605 Daily Raw Counts

Day	Submissions	Minimum	Average	Maximum
1.0000	34.0000	35.0000	76.7000	122.0000
2.0000	35.0000	35.0000	65.5484	131.0000
3.0000	32.0000	26.0000	60.0667	94.0000
4.0000	36.0000	23.0000	59.9697	99.0000
5.0000	33.0000	14.0000	53.2667	98.0000
6.0000	34.0000	11.0000	42.7000	69.0000
7.0000	35.0000	12.0000	31.2258	54.0000
8.0000	33.0000	15.0000	35.7037	58.0000
9.0000	34.0000	13.0000	45.6207	71.0000
10.0000	30.0000	25.0000	63.6667	96.0000
11.0000	33.0000	27.0000	71.9677	103.0000
12.0000	33.0000	42.0000	72.5806	98.0000
13.0000	33.0000	51.0000	71.6000	101.0000
14.0000	34.0000	48.0000	82.6129	117.0000
15.0000	33.0000	26.0000	74.3939	99.0000
16.0000	35.0000	23.0000	45.3438	94.0000
17.0000	33.0000	23.0000	32.2581	60.0000
18.0000	31.0000	22.0000	28.5667	54.0000
19.0000	34.0000	22.0000	39.6452	68.0000
20.0000	35.0000	11.0000	32.9118	57.0000
21.0000	32.0000	11.0000	18.9333	37.0000
22.0000	33.0000	11.0000	17.6000	36.0000
23.0000	35.0000	11.0000	15.3030	37.0000
24.0000	34.0000	11.0000	20.8387	36.0000
25.0000	26.0000	11.0000	31.2083	60.0000
26.0000	32.0000	16.0000	30.7097	54.0000
27.0000	33.0000	11.0000	26.1290	79.0000
28.0000	37.0000	11.0000	29.2941	64.0000
29.0000	35.0000	0.0000	25.4848	57.0000
30.0000	29.0000	13.0000	33.0690	67.0000
31.0000	34.0000	13.0000	31.5455	49.0000

3 Error Tables

Data are for the month of May 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	23.2604	22.7292	23.7915	8.4000	8.7000
2010.06	18.7737	18.2716	19.2758	11.0000	13.6000
2010.07	20.5929	20.1393	21.0465	15.2000	16.1000
2010.08	19.6394	19.1606	20.1183	18.3000	19.6000
2010.09	23.6492	23.1423	24.1562	22.8000	25.2000
2010.10	22.7129	22.2227	23.2030	21.0000	23.5000
2010.11	24.0288	23.4847	24.5730	20.9000	21.6000
2010.12	23.2529	22.5842	23.9216	13.9000	14.5000
2011.01	74.3126	72.6440	75.9812	17.7000	18.7000
2011.02	64.9747	63.5057	66.4438	29.1000	29.6000
2011.03	70.6291	69.1705	72.0878	48.0000	55.8000
2011.04	75.7548	74.1318	77.3778	47.3000	54.4000
2011.05	79.0353	77.4517	80.6188	37.3000	41.5000
2011.06	67.2702	65.8551	68.6853	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	71.9067	70.3286	73.4847	41.5000	43.8000
2011.08	71.9011	70.4910	73.3112	42.4000	50.5000
2011.09	83.8450	82.7490	84.9411	73.8000	78.0000
2011.10	79.9995	78.6322	81.3668	78.9000	88.0000
2011.11	83.2222	81.4476	84.9968	84.6000	96.7000
2011.12	78.1892	76.4824	79.8960	65.8000	73.0000
2012.01	76.2687	74.7570	77.7804	55.8000	58.2000
2012.02	64.5013	63.1370	65.8655	29.2000	33.1000
2012.03	73.0337	71.7314	74.3359	53.1000	64.1000
2012.04	75.4172	73.2112	77.6232	51.4000	55.2000
2012.05	82.9837	81.5367	84.4308	61.8000	69.0000
2012.06	70.4017	69.1542	71.6492	59.7000	64.5000
2012.07	76.2170	74.9406	77.4934	64.2000	51.3000
2012.08	72.6097	71.3975	73.8219	57.7000	63.1000
2012.09	84.5773	83.1334	86.0213	57.7000	61.5000
2012.10	82.1975	80.6441	83.7509	48.3000	53.3000
2012.11	86.6166	84.8796	88.3537	56.7000	61.4000
2012.12	79.0765	77.4029	80.7502	37.4000	40.8000
2013.01	86.1727	84.5482	87.7973	63.8000	62.9000
2013.02	74.5711	73.1158	76.0263	37.8000	38.0000
2013.03	80.2828	78.7719	81.7937	50.6000	57.9000
2013.04	88.0392	86.5528	89.5255	70.6000	72.4000
2013.05	90.7974	89.2282	92.3666	77.4000	78.7000
2013.06	77.6722	76.2879	79.0566	51.0000	52.5000
2013.07	81.9825	80.7016	83.2635	57.0000	57.0000
2013.08	80.1904	78.9348	81.4460	60.0000	66.0000
2013.09	92.5492	90.9462	94.1521	34.6000	36.9000
2013.10	88.4329	86.8584	90.0074	74.5000	85.6000
2013.11	93.3460	91.3951	95.2969	73.9000	77.6000
2013.12	87.4791	85.7130	89.2452	77.8000	90.3000
2014.01	102.4798	100.3193	104.6403	77.4000	82.0000
2014.02	88.6710	86.9815	90.3605	93.9000	102.8000
2014.03	98.7769	97.1085	100.4453	80.9000	92.2000
2014.04	106.9937	105.1710	108.8163	76.9000	84.7000
2014.05	109.9971	108.2372	111.7570	72.3000	75.2000
2014.06	94.0882	92.5564	95.6199	67.2000	71.0000
2014.07	100.7292	99.0878	102.3705	72.5000	72.5000
2014.08	98.0051	96.5248	99.4854	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	114.1127	112.2822	115.9432	83.2000	87.6000
2014.10	108.8926	107.0652	110.7201	59.5000	60.6000
2014.11	115.4739	113.2812	117.6667	65.8000	71.1000
2014.12	105.5902	103.3126	107.8677	75.8000	78.0000
2015.01	62.2757	61.0764	63.4750	65.9000	67.0000
2015.02	53.8791	52.6299	55.1284	42.4000	44.8000
2015.03	58.9503	57.8729	60.0277	38.0000	38.4000
2015.04	64.6693	63.5267	65.8119	49.0000	54.4000
2015.05	66.1912	65.1227	67.2598	56.3000	58.8000
2015.06	56.9969	56.0321	57.9618	50.2000	68.3000
2015.07	59.6751	58.6596	60.6906	47.9000	65.8000
2015.08	59.4308	58.4601	60.4015	39.5000	57.2000
2015.09	69.0523	67.9291	70.1754	49.2000	72.1000
2015.10	65.8336	64.7093	66.9579	39.3000	48.3000
2015.11	70.3394	69.3850	71.2938	39.6000	55.9000
2015.12	63.7552	62.4577	65.0528	36.4000	44.8000
2016.01	44.1062	43.3240	44.8884	33.7000	43.3000
2016.02	37.3491	36.6084	38.0898	38.3000	46.8000
2016.03	40.6998	39.9556	41.4440	30.5000	38.9000
2016.04	43.9018	43.1216	44.6820	26.6000	30.9000
2016.05	46.0061	45.2168	46.7953	33.7000	48.4000

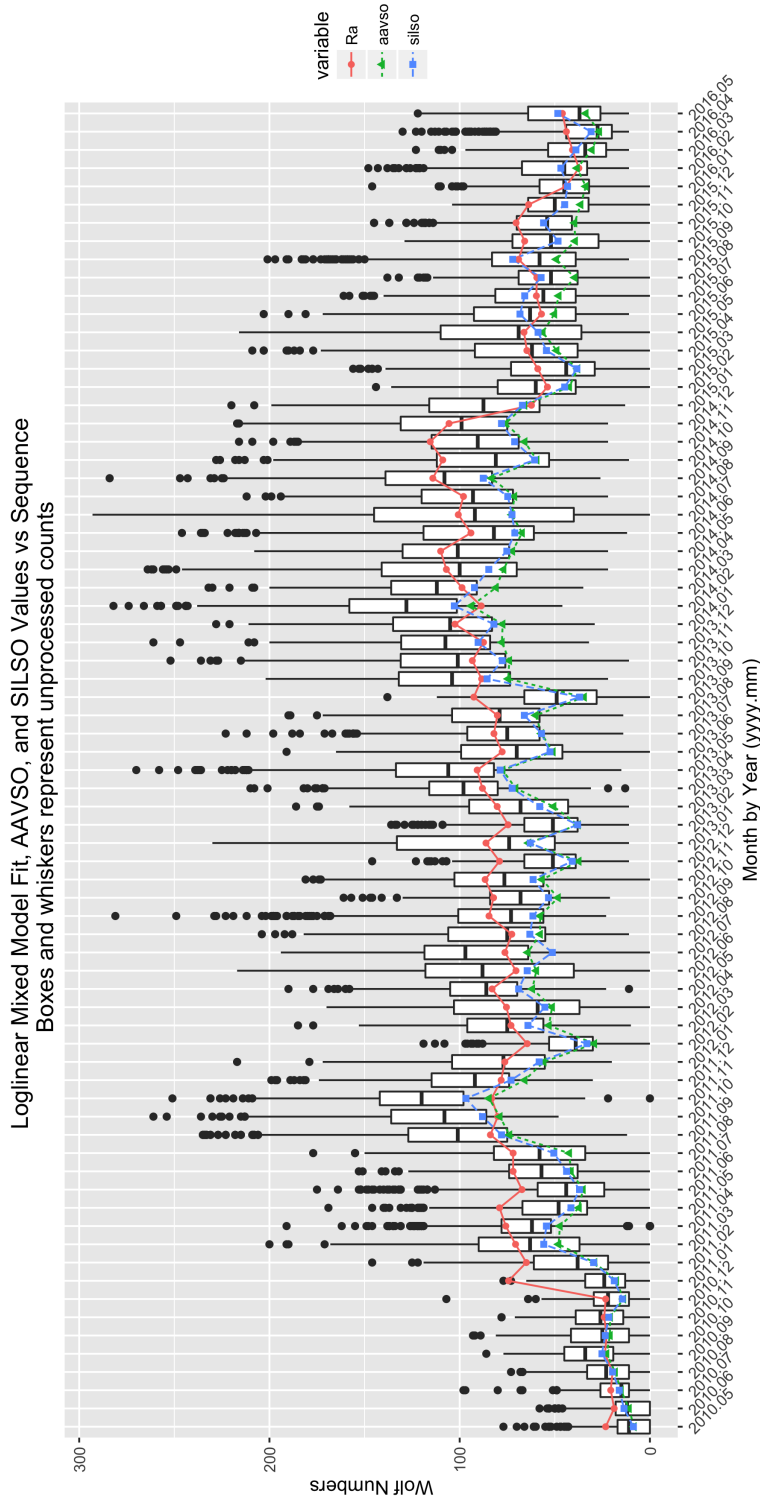


Figure 2: GLMM fitted data for R_a . AAVSO data: <https://www.aavso.org/category/tags/solar-bulletin>. SILSO data: WDC-SILSO, Royal Observatory of Belgium, Brussels

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: 201605 Parameter Estimates

	Estimate	Std. Error	t-value	$\Pr(> t)$
(Intercept)	3.1889	0.0445	71.6976	0.0000
seeF	-0.1840	0.0072	-25.4181	0.0000
seeG	-0.0970	0.0063	-15.3854	0.0000
seeP	-0.2940	0.0106	-27.7251	0.0000
r1000B	-0.0645	0.0832	-0.7749	0.4384
r1500C	0.0279	0.1272	0.2195	0.8263
r2000D	0.0745	0.1551	0.4800	0.6312
r2500E	-0.0001	0.1054	-0.0010	0.9992
r3000F	0.0660	0.1026	0.6435	0.5199
r3500G	0.1189	0.1535	0.7743	0.4388
r5000H	-0.1142	0.2124	-0.5380	0.5906
silsoy	0.1175	0.0740	1.5887	0.1121
year2011	1.2147	0.0151	80.2570	0.0000
year2012	1.2301	0.0151	81.5359	0.0000
year2013	1.3280	0.0150	88.2754	0.0000
year2014	1.5171	0.0149	101.5146	0.0000
year2015	1.0167	0.0153	66.3151	0.0000
year2016	0.6507	0.0185	35.2017	0.0000
mon2	-0.1537	0.0119	-12.9258	0.0000
mon3	-0.0641	0.0109	-5.8614	0.0000
mon4	0.0197	0.0110	1.7910	0.0733
mon5	0.0497	0.0104	4.7739	0.0000
mon6	-0.1154	0.0112	-10.3265	0.0000
mon7	-0.0531	0.0108	-4.9314	0.0000
mon8	-0.0700	0.0107	-6.5665	0.0000
mon9	0.0842	0.0102	8.2523	0.0000
mon10	0.0429	0.0108	3.9747	0.0001
mon11	0.1054	0.0110	9.5967	0.0000
mon12	0.0266	0.0117	2.2793	0.0227

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.

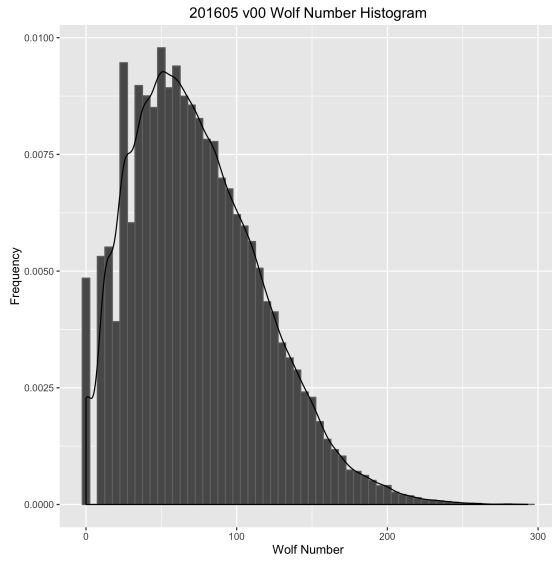
5 Supporting Information

Table 4: 201605 Summary of Sunspot Numbers

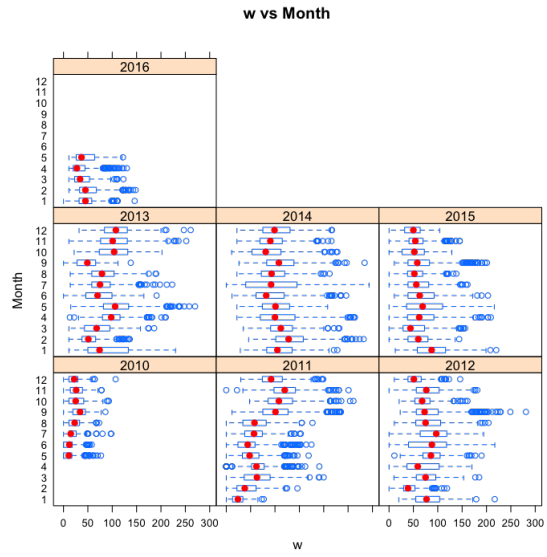
obs	jd	year	mon	day
ARAG : 2168	Min. :1721096	Min. :2010	Min. : 1.000	Min. : 1.00
CHAG : 1988	1st Qu.:2455944	1st Qu.:2012	1st Qu.: 4.000	1st Qu.: 8.00
BRAB : 1953	Median :2456466	Median :2013	Median : 7.000	Median :16.00
BROB : 1747	Mean :2456115	Mean :2013	Mean : 6.589	Mean :15.74
DUBF : 1648	3rd Qu.:2456969	3rd Qu.:2014	3rd Qu.: 9.000	3rd Qu.:23.00
HOWR : 1616	Max. :2457540	Max. :2016	Max. :12.000	Max. :31.00
(Other):39572				

Table 5: Summary of Sunspot Numbers

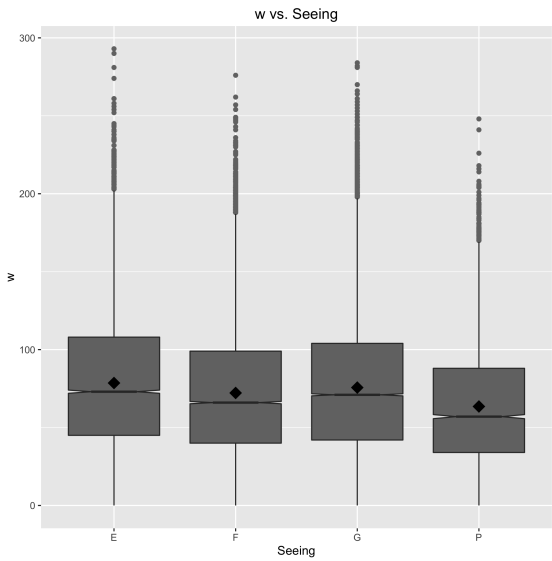
see	g	s	w	r	silso
E: 9189	Min. : 0.000	Min. : 0.00	Min. : 0.0	0000A :21692	n:34143
F:15594	1st Qu.: 3.000	1st Qu.: 11.00	1st Qu.: 41.0	3000F : 8259	y:16549
G:21687	Median : 4.000	Median : 23.00	Median : 69.0	2500E : 6573	
P: 4222	Mean : 4.554	Mean : 28.56	Mean : 74.1	3500G : 3941	
	3rd Qu.: 6.000	3rd Qu.: 41.00	3rd Qu.:102.0	1000B : 3641	
	Max. :18.000	Max. :204.00	Max. :293.0	1500C : 2938	
				(Other): 3648	



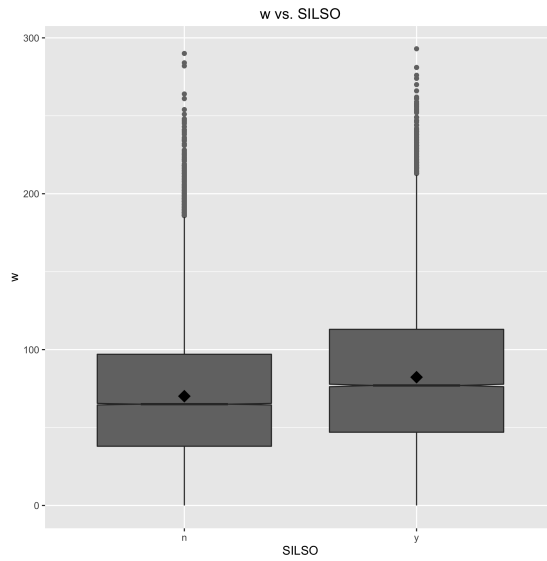
(a) Observed sunspot count histogram.



(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:

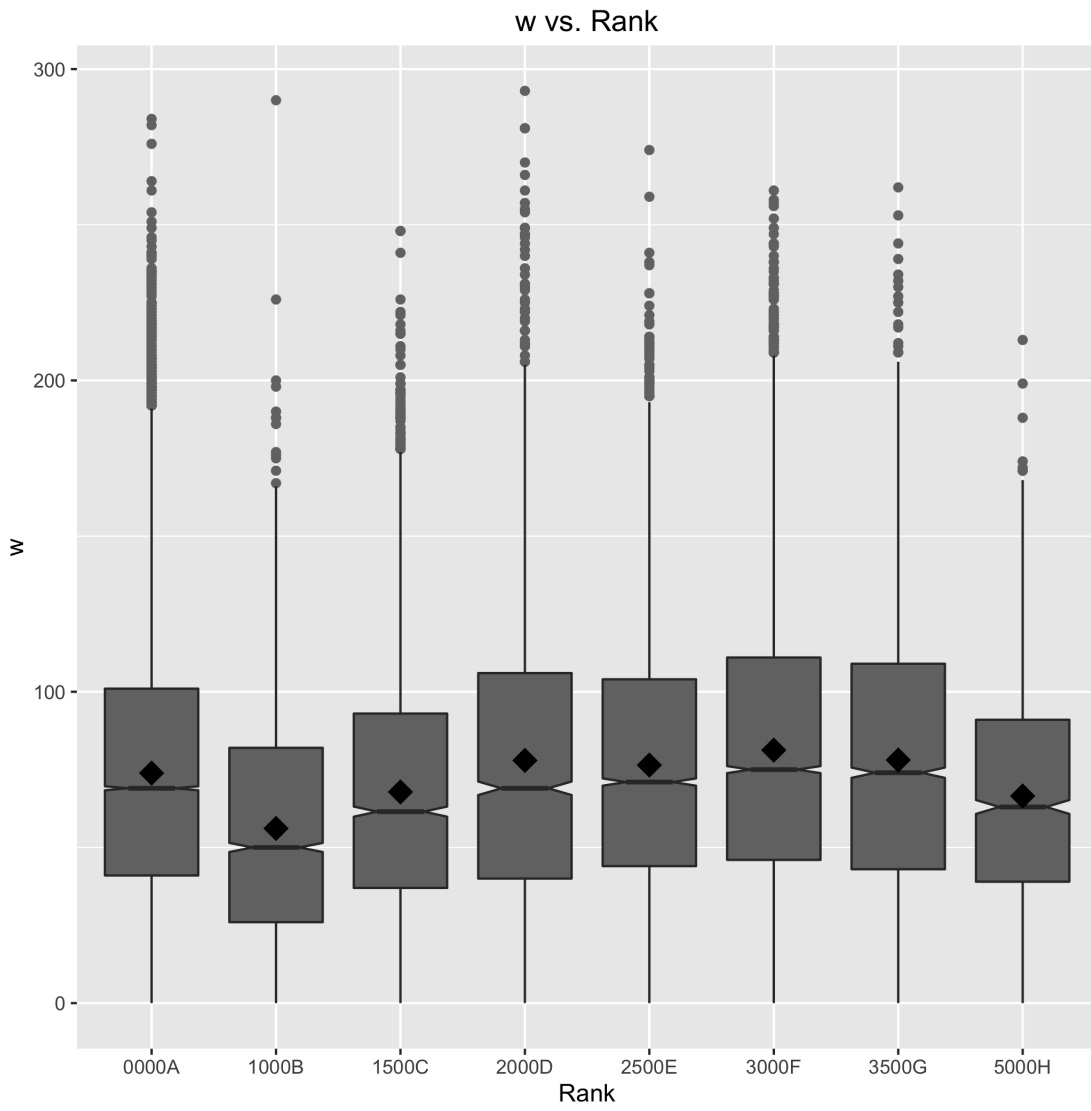


Figure 4: Box plot of sunspot count by rank.