

Monthly Report (00)

2016.03 Data Set

Thursday 14th April, 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.

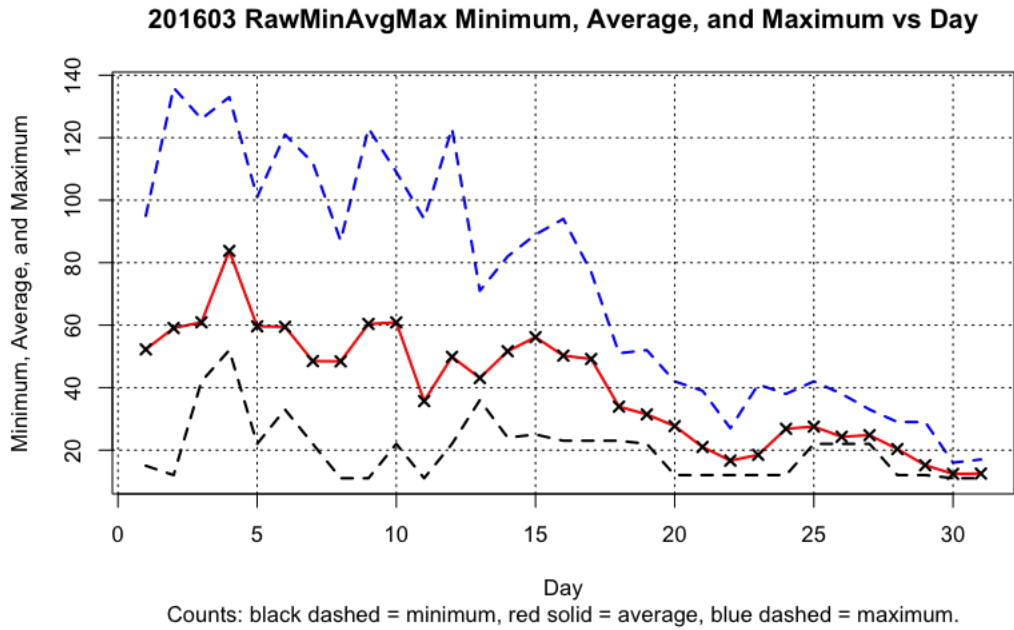


Figure 1: Raw average sunspot count by day of the month.

Table 1: 201603 Daily Raw Counts

Day	Submissions	Minimum	Average	Maximum
1.0000	29.0000	15.0000	52.2500	95.0000
2.0000	34.0000	12.0000	59.0714	136.0000
3.0000	31.0000	42.0000	60.9231	126.0000
4.0000	25.0000	52.0000	83.7727	133.0000
5.0000	31.0000	22.0000	59.6296	101.0000
6.0000	35.0000	33.0000	59.4643	121.0000
7.0000	30.0000	22.0000	48.5000	112.0000
8.0000	36.0000	11.0000	48.4194	87.0000
9.0000	28.0000	11.0000	60.3750	123.0000
10.0000	20.0000	22.0000	60.8889	109.0000
11.0000	33.0000	11.0000	35.7000	94.0000
12.0000	37.0000	22.0000	49.8571	123.0000
13.0000	32.0000	36.0000	43.0769	71.0000
14.0000	26.0000	24.0000	51.6522	82.0000
15.0000	28.0000	25.0000	56.1538	89.0000
16.0000	30.0000	23.0000	50.2400	94.0000
17.0000	37.0000	23.0000	49.1724	77.0000
18.0000	27.0000	23.0000	33.9200	51.0000
19.0000	28.0000	22.0000	31.4348	52.0000
20.0000	33.0000	12.0000	27.7037	42.0000
21.0000	35.0000	12.0000	21.0000	39.0000
22.0000	33.0000	12.0000	16.6452	27.0000
23.0000	28.0000	12.0000	18.4815	41.0000
24.0000	32.0000	12.0000	26.8333	38.0000
25.0000	24.0000	22.0000	27.5217	42.0000
26.0000	42.0000	22.0000	24.2812	38.0000
27.0000	32.0000	22.0000	24.7931	33.0000
28.0000	33.0000	12.0000	20.3448	29.0000
29.0000	34.0000	12.0000	15.1667	29.0000
30.0000	31.0000	11.0000	12.3929	16.0000
31.0000	29.0000	11.0000	12.5000	17.0000

3 Error Tables

Data are for the month of March 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	23.3258	22.7931	23.8585	8.4000	8.7000
2010.06	18.7723	18.2703	19.2743	11.0000	13.6000
2010.07	20.5838	20.1304	21.0371	15.2000	16.1000
2010.08	19.6353	19.1574	20.1132	18.3000	19.6000
2010.09	23.6372	23.1306	24.1437	22.8000	25.2000
2010.10	22.6962	22.2061	23.1863	21.0000	23.5000
2010.11	24.0164	23.4723	24.5604	20.9000	21.6000
2010.12	23.2498	22.5817	23.9179	13.9000	14.5000
2011.01	73.4148	71.7664	75.0632	17.7000	18.7000
2011.02	64.1833	62.7318	65.6348	29.1000	29.6000
2011.03	69.8185	68.3778	71.2593	48.0000	55.8000
2011.04	77.7576	76.0898	79.4254	47.3000	54.4000
2011.05	79.2669	77.6770	80.8567	37.3000	41.5000
2011.06	67.2552	65.8380	68.6724	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	71.8866	70.3076	73.4656	41.5000	43.8000
2011.08	71.8802	70.4705	73.2899	42.4000	50.5000
2011.09	83.8507	82.7570	84.9443	73.8000	78.0000
2011.10	80.0002	78.6361	81.3642	78.9000	88.0000
2011.11	83.2495	81.4792	85.0198	84.6000	96.7000
2011.12	78.2034	76.5005	79.9063	65.8000	73.0000
2012.01	75.4965	74.0002	76.9928	55.8000	58.2000
2012.02	63.8369	62.4899	65.1838	29.2000	33.1000
2012.03	72.3456	71.0582	73.6330	53.1000	64.1000
2012.04	77.5336	75.2755	79.7917	51.4000	55.2000
2012.05	83.3875	81.9351	84.8398	61.8000	69.0000
2012.06	70.5124	69.2640	71.7607	59.7000	64.5000
2012.07	76.3506	75.0736	77.6276	64.2000	51.3000
2012.08	72.7458	71.5331	73.9585	57.7000	63.1000
2012.09	84.7208	83.2750	86.1666	57.7000	61.5000
2012.10	82.3241	80.7693	83.8789	48.3000	53.3000
2012.11	86.7581	85.0192	88.4970	56.7000	61.4000
2012.12	79.2001	77.5246	80.8755	37.4000	40.8000
2013.01	85.1108	83.5055	86.7160	63.8000	62.9000
2013.02	73.6170	72.1829	75.0511	37.8000	38.0000
2013.03	79.3099	77.8200	80.7998	50.6000	57.9000
2013.04	90.3383	88.8136	91.8630	70.6000	72.4000
2013.05	91.0407	89.4684	92.6131	77.4000	78.7000
2013.06	77.6270	76.2443	79.0098	51.0000	52.5000
2013.07	81.9356	80.6562	83.2150	57.0000	57.0000
2013.08	80.1439	78.8898	81.3981	60.0000	66.0000
2013.09	92.4959	90.8958	94.0961	34.6000	36.9000
2013.10	88.4072	86.8351	89.9793	74.5000	85.6000
2013.11	93.2907	91.3428	95.2386	73.9000	77.6000
2013.12	87.4440	85.6798	89.2081	77.8000	90.3000
2014.01	101.2546	99.1252	103.3840	77.4000	82.0000
2014.02	87.5885	85.9239	89.2530	93.9000	102.8000
2014.03	97.6034	95.9565	99.2502	80.9000	92.2000
2014.04	109.7985	107.9284	111.6687	76.9000	84.7000
2014.05	110.3142	108.5512	112.0773	72.3000	75.2000
2014.06	94.0474	92.5163	95.5784	67.2000	71.0000
2014.07	100.6750	99.0364	102.3136	72.5000	72.5000
2014.08	97.9502	96.4710	99.4294	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	114.0456	112.2178	115.8735	83.2000	87.6000
2014.10	108.8454	107.0194	110.6714	59.5000	60.6000
2014.11	115.4239	113.2365	117.6113	65.8000	71.1000
2014.12	105.5256	103.2539	107.7972	75.8000	78.0000
2015.01	61.5150	60.3333	62.6968	65.9000	67.0000
2015.02	53.2061	51.9772	54.4350	42.4000	44.8000
2015.03	58.2448	57.1822	59.3075	38.0000	38.4000
2015.04	66.3628	65.1913	67.5343	49.0000	54.4000
2015.05	66.3677	65.2963	67.4390	56.3000	58.8000
2015.06	56.9748	56.0109	57.9388	50.2000	68.3000
2015.07	59.6531	58.6392	60.6670	47.9000	65.8000
2015.08	59.3977	58.4273	60.3682	39.5000	57.2000
2015.09	69.0102	67.8878	70.1325	49.2000	72.1000
2015.10	65.7893	64.6651	66.9136	39.3000	48.3000
2015.11	70.2998	69.3462	71.2534	39.6000	55.9000
2015.12	63.7030	62.4030	65.0030	36.4000	44.8000
2016.01	49.2835	48.4104	50.1565	33.7000	43.3000
2016.02	41.7038	40.8776	42.5300	38.3000	46.8000
2016.03	45.4969	44.6671	46.3266	30.5000	38.9000

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

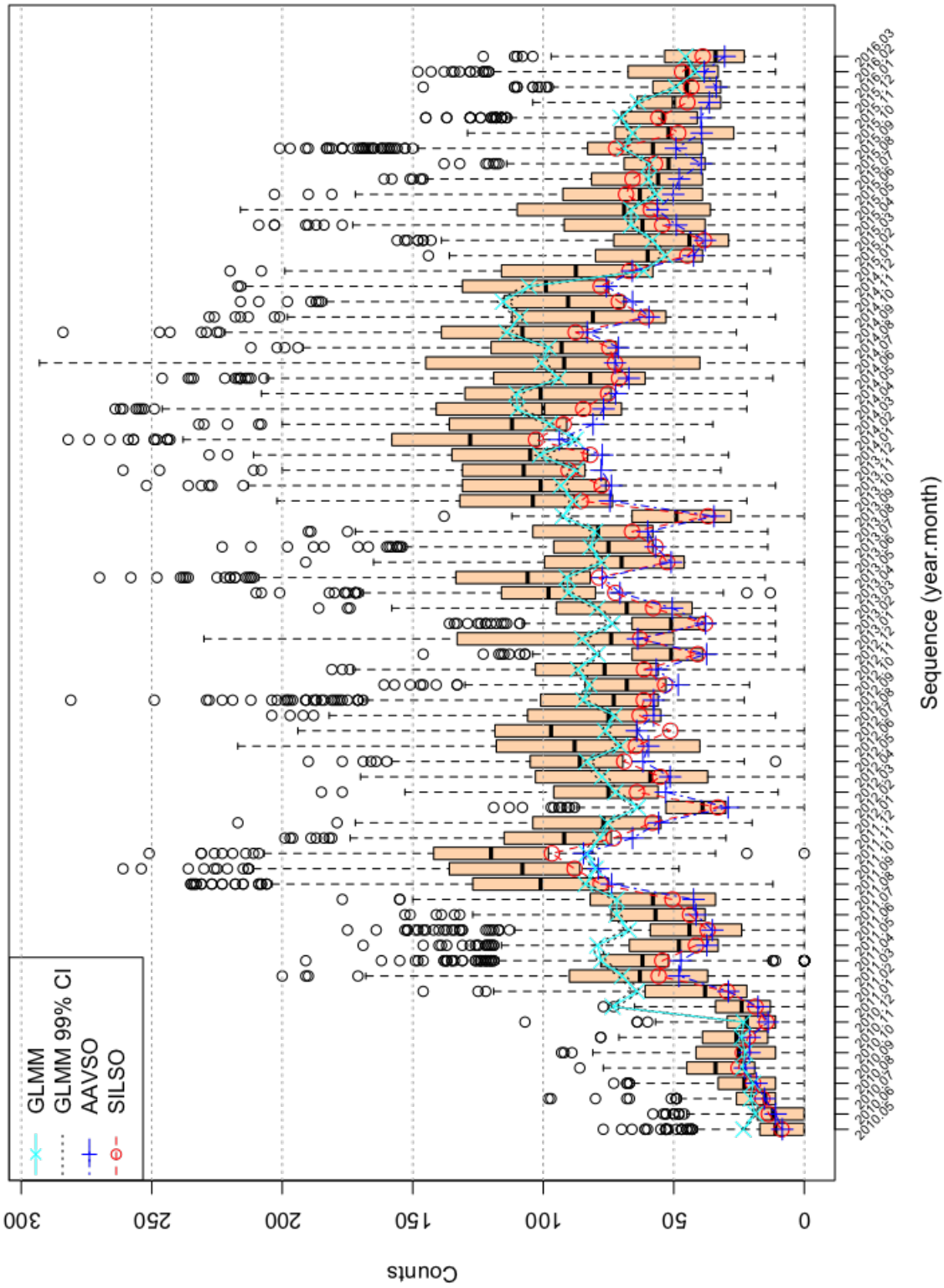


Figure 2: GLMM fitted data for R_a . AAVSO data: <https://www.aavso.org/category/tags/solar-bulletin>. SILSO data: WDC-SILSO, Royal Observatory of Belgium, Brussels

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: 201603 Parameter Estimates

	Estimate	Std. Error	t-value	$\Pr(> t)$
(Intercept)	3.1783	0.0444	71.5238	0.0000
seeF	-0.1847	0.0073	-25.2453	0.0000
seeG	-0.0968	0.0064	-15.1914	0.0000
seeP	-0.2962	0.0107	-27.6334	0.0000
r1000B	-0.0662	0.0830	-0.7972	0.4253
r1500C	0.0266	0.1269	0.2098	0.8338
r2000D	0.0745	0.1548	0.4812	0.6304
r2500E	-0.0009	0.1051	-0.0086	0.9932
r3000F	0.0630	0.1024	0.6151	0.5385
r3500G	0.1182	0.1532	0.7720	0.4401
r5000H	-0.1168	0.2118	-0.5515	0.5813
silsoy	0.1180	0.0738	1.5984	0.1100
year2011	1.2147	0.0152	79.9891	0.0000
year2012	1.2319	0.0151	81.3825	0.0000
year2013	1.3275	0.0151	87.9475	0.0000
year2014	1.5168	0.0150	101.1508	0.0000
year2015	1.0164	0.0154	66.0738	0.0000
year2016	0.7739	0.0212	36.5399	0.0000
mon2	-0.1541	0.0119	-12.9156	0.0000
mon3	-0.0640	0.0110	-5.8353	0.0000
mon4	0.0577	0.0114	5.0710	0.0000
mon5	0.0645	0.0107	6.0007	0.0000
mon6	-0.1040	0.0113	-9.2233	0.0000
mon7	-0.0416	0.0109	-3.8307	0.0001
mon8	-0.0585	0.0108	-5.4396	0.0000
mon9	0.0957	0.0103	9.2917	0.0000
mon10	0.0546	0.0109	5.0149	0.0000
mon11	0.1172	0.0111	10.5858	0.0000
mon12	0.0385	0.0118	3.2735	0.0011

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.

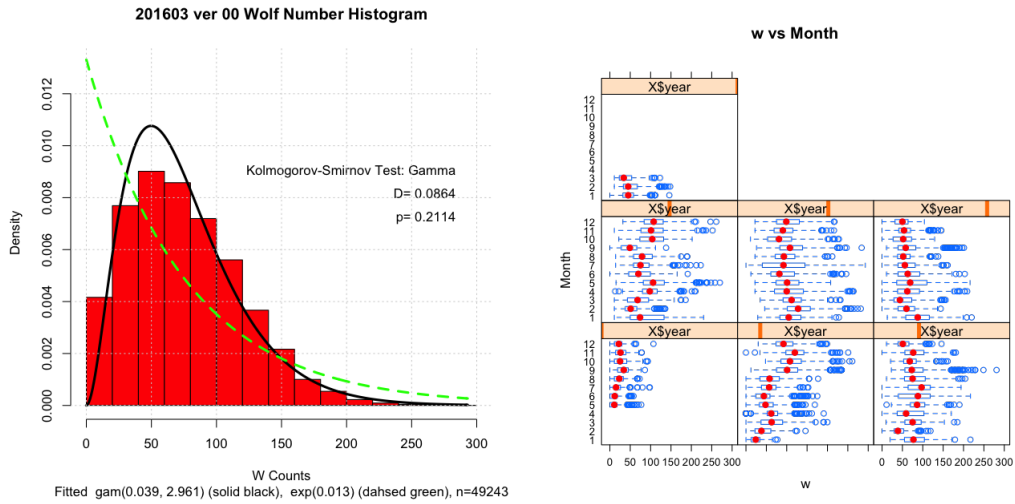
5 Supporting Information

Table 4: 201603 Summary of Sunspot Numbers

obs	jd	year	mon	day
ARAG : 2108	Min. :1721096	Min. :2010	Min. : 1.000	Min. : 1.00
CHAG : 1928	1st Qu.:2455927	1st Qu.:2011	1st Qu.: 4.000	1st Qu.: 8.00
BRAB : 1894	Median :2456440	Median :2013	Median : 7.000	Median :16.00
BROB : 1694	Mean :2456074	Mean :2013	Mean : 6.651	Mean :15.74
DUBF : 1590	3rd Qu.:2456923	3rd Qu.:2014	3rd Qu.: 9.000	3rd Qu.:23.00
HOWR : 1569	Max. :2457479	Max. :2016	Max. :12.000	Max. :31.00
(Other):38460				

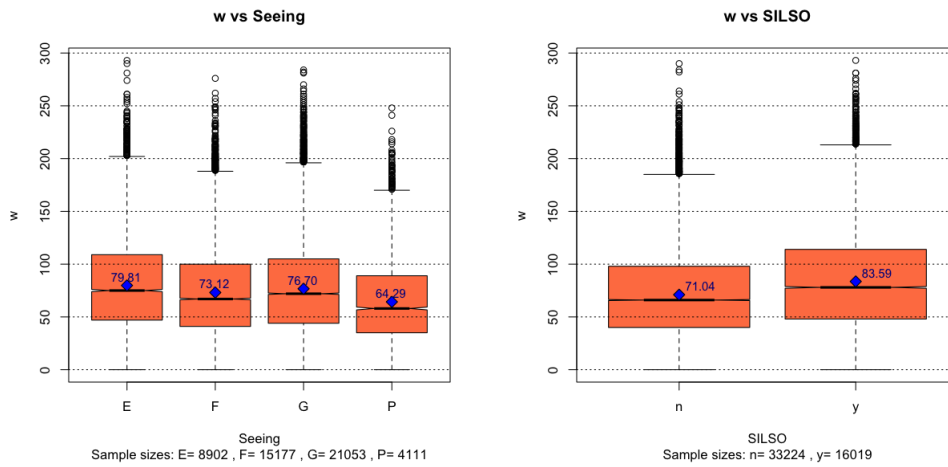
Table 5: Summary of Sunspot Numbers

see	g	s	w	r	silso
E: 8902	Min. : 0.000	Min. : 0.00	Min. : 0.00	0000A :21063	n:33224
F:15177	1st Qu.: 3.000	1st Qu.: 11.00	1st Qu.: 42.00	3000F : 8012	y:16019
G:21053	Median : 4.000	Median : 23.00	Median : 70.00	2500E : 6385	
P: 4111	Mean : 4.611	Mean : 29.01	Mean : 75.12	3500G : 3822	
	3rd Qu.: 6.000	3rd Qu.: 41.00	3rd Qu.:103.00	1000B : 3552	
	Max. :18.000	Max. :204.00	Max. :293.00	1500C : 2901	
				(Other): 3508	



(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:

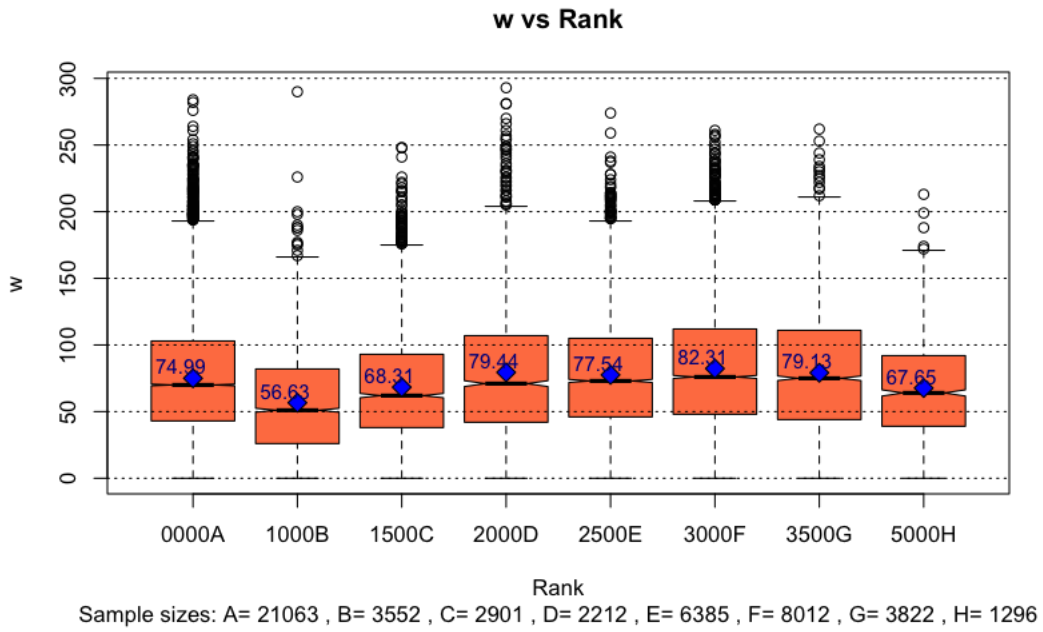


Figure 4: Box plot of sunspot count by rank.