Monthly Report (00)

2015.06 Data Set

Sunday 19th July, 2015

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.

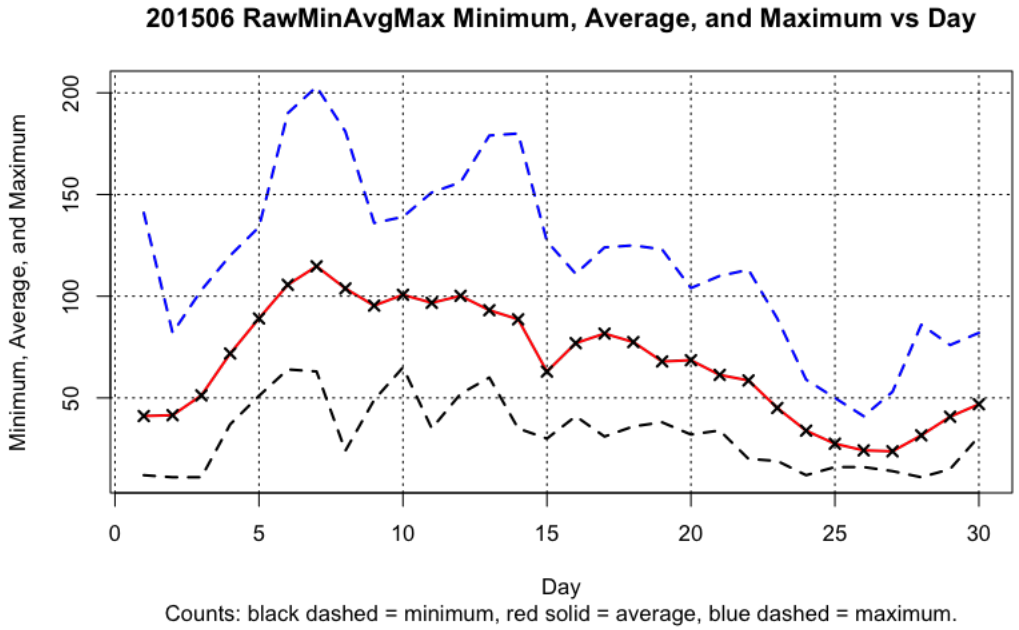


Figure 1: Raw average sunspot count by day of the month.

Table 1: 201506 Daily Raw Counts

Day	Submissions	Minimum	Average	Maximum
1.0000	30.0000	12.0000	41.1034	141.0000
2.0000	31.0000	11.0000	41.5000	82.0000
3.0000	30.0000	11.0000	51.2414	103.0000
4.0000	40.0000	37.0000	71.8333	120.0000
5.0000	33.0000	51.0000	89.0667	134.0000
6.0000	41.0000	64.0000	105.6286	190.0000
7.0000	41.0000	63.0000	114.7143	203.0000
8.0000	35.0000	24.0000	103.8125	181.0000
9.0000	29.0000	49.0000	95.3846	136.0000
10.0000	32.0000	65.0000	100.6923	139.0000
11.0000	32.0000	35.0000	96.7407	151.0000
12.0000	32.0000	52.0000	100.1786	156.0000
13.0000	28.0000	60.0000	93.1200	179.0000
14.0000	26.0000	35.0000	88.6364	180.0000
15.0000	27.0000	30.0000	62.9600	127.0000
16.0000	32.0000	41.0000	76.9333	111.0000
17.0000	32.0000	31.0000	81.6296	124.0000
18.0000	35.0000	36.0000	77.4545	125.0000
19.0000	32.0000	38.0000	67.9286	123.0000
20.0000	37.0000	32.0000	68.5312	104.0000
21.0000	36.0000	34.0000	61.3030	110.0000
22.0000	40.0000	20.0000	58.5882	113.0000
23.0000	36.0000	19.0000	45.0588	89.0000
24.0000	35.0000	12.0000	33.8824	59.0000
25.0000	37.0000	16.0000	27.4848	50.0000
26.0000	34.0000	16.0000	24.2000	41.0000
27.0000	30.0000	14.0000	23.7500	53.0000
28.0000	30.0000	11.0000	31.5926	86.0000
29.0000	42.0000	15.0000	40.7568	76.0000
30.0000	41.0000	31.0000	46.8919	82.0000

3 Error Tables

Data are for the month of June 2015. 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 25^{th} through the 75^{th} quartiles. The lower and upper whiskers extend 1.5 times the IQR below the 25^{th} quartile, and 1.5 times the IQR above the 75^{th} 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
2010.05	22.4942	21.9845	23.0039
2010.06	18.1024	17.6230	18.5818
2010.07	20.1680	19.7257	20.6103
2010.08	19.6797	19.2052	20.1543
2010.09	23.5286	23.0315	24.0257
2010.10	23.1206	22.6260	23.6152
2010.11	25.2346	24.6688	25.8003
2010.12	23.7638	23.0856	24.4419
2011.01	71.7366	70.1293	73.3439
2011.02	60.7759	59.4048	62.1470
2011.03	69.0594	67.6405	70.4783
2011.04	75.1448	73.5363	76.7534
2011.05	77.1715	75.6249	78.7181
2011.06	65.5126	64.1363	66.8889

Continued on next page

Table 2: Year Month (ym) Relative Sunspot Numbers with 99% CI

ym	Ra	lci99	uci99
2011.07	71.1322	69.5683	72.6961
2011.08	72.6109	71.1872	74.0345
2011.09	84.1383	83.0371	85.2394
2011.10	82.3935	80.9910	83.7959
2011.11	88.4837	86.6238	90.3437
2011.12	80.8298	79.0821	82.5774
2012.01	73.9114	72.4583	75.3646
2012.02	60.5770	59.3048	61.8492
2012.03	71.6803	70.4074	72.9532
2012.04	75.3609	73.2517	77.4702
2012.05	81.1938	79.7797	82.6078
2012.06	68.7247	67.5103	69.9391
2012.07	75.5986	74.3329	76.8643
2012.08	73.5802	72.3567	74.8038
2012.09	85.1931	83.7452	86.6410
2012.10	84.9133	83.3150	86.5115
2012.11	92.2927	90.4534	94.1320
2012.12	81.9828	80.2523	83.7133
2013.01	83.5998	82.0319	85.1678
2013.02	69.8610	68.4974	71.2247
2013.03	78.6505	77.1675	80.1336
2013.04	87.7131	86.2416	89.1846
2013.05	89.0159	87.4896	90.5422
2013.06	75.8368	74.4917	77.1820
2013.07	81.3194	80.0544	82.5843
2013.08	81.2290	79.9631	82.4949
2013.09	93.2450	91.6421	94.8479
2013.10	91.5838	89.9686	93.1990
2013.11	99.5271	97.4673	101.5868
2013.12	90.6914	88.8667	92.5161
2014.01	99.3610	97.2760	101.4461
2014.02	83.4126	81.8385	84.9866
2014.03	96.9100	95.2725	98.5476
2014.04	106.6254	104.8233	108.4275
2014.05	107.8264	106.1079	109.5450
2014.06	91.9715	90.4820	93.4610
2014.07	100.0186	98.3947	101.6426
2014.08	99.2615	97.7730	100.7499

Continued on next page

Table 2: Year Month (ym) Relative Sunspot Numbers with 99% CI

ym	Ra	lci99	uci99
2014.09	115.1041	113.2785	116.9296
2014.10	112.5932	110.7161	114.4703
2014.11	123.0459	120.7106	125.3811
2014.12	109.6005	107.2559	111.9451
2015.01	70.6354	69.2705	72.0002
2015.02	59.2016	57.8350	60.5682
2015.03	67.6755	66.4411	68.9099
2015.04	75.2669	73.9363	76.5975
2015.05	75.8006	74.5887	77.0124
2015.06	65.0860	63.9758	66.1962

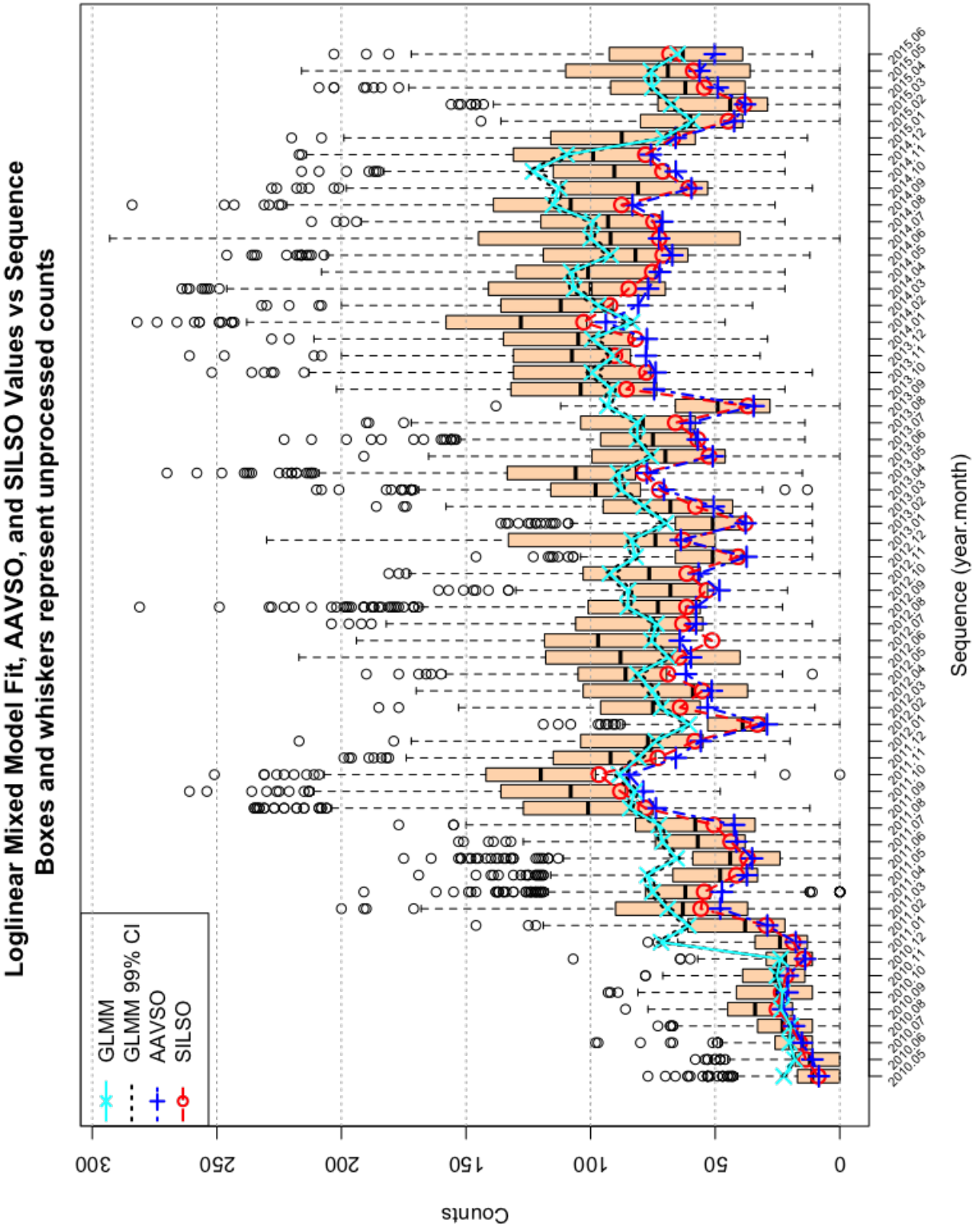


Figure 2: GLMM fitted data for R_a.

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: 201506 Parameter Estimates

	Estimate	Std. Error	t-value	Pr(>|t|)
(Intercept)	3.1361	0.0447	70.2014	0.0000
seeF	-0.1837	0.0078	-23.6054	0.0000
seeG	-0.0941	0.0068	-13.8745	0.0000
seeP	-0.2943	0.0114	-25.7411	0.0000
r1000B	-0.0503	0.0831	-0.6053	0.5450
r1500C	0.0421	0.1270	0.3317	0.7401
r2000D	0.0854	0.1548	0.5518	0.5811
r2500E	0.0010	0.1052	0.0097	0.9922
r3000F	0.0623	0.1024	0.6086	0.5428
r3500G	0.1218	0.1532	0.7953	0.4265
r5000H	-0.1053	0.2119	-0.4969	0.6193
silsoy	0.1238	0.0738	1.6771	0.0935
year2011	1.2263	0.0154	79.5453	0.0000
year2012	1.2449	0.0154	81.0020	0.0000
year2013	1.3428	0.0153	87.6053	0.0000
year2014	1.5329	0.0152	100.6462	0.0000
year2015	1.1901	0.0170	70.0693	0.0000
mon2	-0.1864	0.0129	-14.4828	0.0000
mon3	-0.0531	0.0117	-4.5481	0.0000
mon4	0.0452	0.0117	3.8545	0.0001
mon5	0.0591	0.0111	5.3191	0.0000
mon6	-0.1094	0.0116	-9.4042	0.0000
mon7	-0.0309	0.0117	-2.6340	0.0084
mon8	-0.0266	0.0115	-2.3023	0.0213
mon9	0.1216	0.0110	11.0524	0.0000
mon10	0.1066	0.0116	9.1627	0.0000
mon11	0.1996	0.0121	16.4840	0.0000
mon12	0.0944	0.0126	7.4963	0.0000

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.

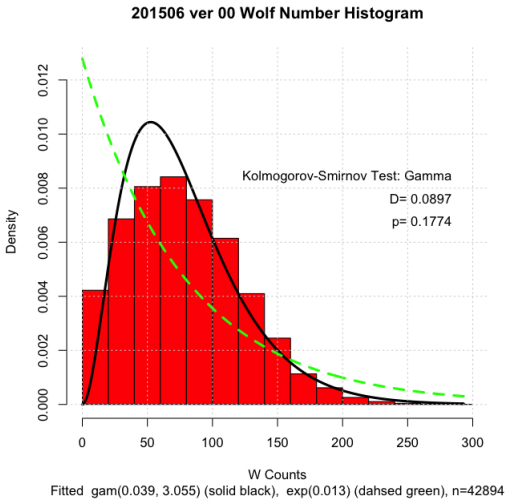
5 Supporting Information

Table 4: 201506 Summary of Sunspot Numbers

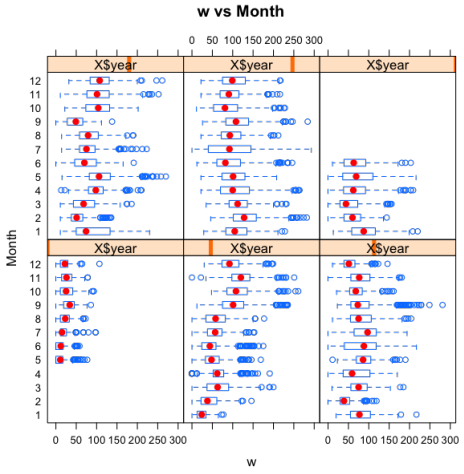
obs	jd	year	mon	day
ARAG : 1823	Min. :1721096	Min. :2010	Min. : 1.000	Min. : 1.00
CHAG : 1647	1st Qu.:2455848	1st Qu.:2011	1st Qu.: 4.000	1st Qu.: 8.00
BRAB : 1615	Median :2456306	Median :2013	Median : 7.000	Median :16.00
BROB : 1474	Mean :2455888	Mean :2013	Mean : 6.534	Mean :15.73
DUBF : 1366	3rd Qu.:2456753	3rd Qu.:2014	3rd Qu.: 9.000	3rd Qu.:23.00
FUJK : 1325	Max. :2457204	Max. :2015	Max. :12.000	Max. :31.00
(Other):33644				

Table 5: Summary of Sunspot Numbers

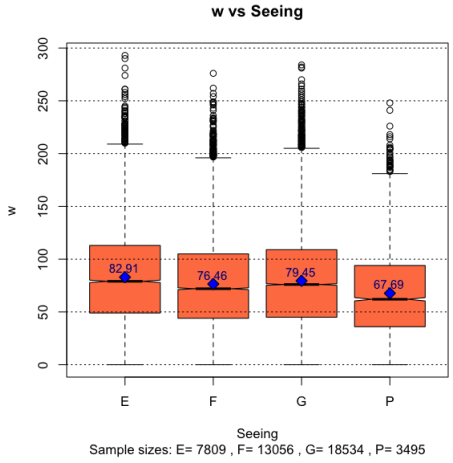
see	g	s	w	r	silso
E: 7809	Min. : 0.000	Min. : 0.00	Min. : 0.00	0000A :18556	n:29106
F:13056	1st Qu.: 3.000	1st Qu.: 12.00	1st Qu.: 45.00	3000F : 6856	y:13788
G:18534	Median : 5.000	Median : 25.00	Median : 74.00	2500E : 5551	
P: 3495	Mean : 4.797	Mean : 30.24	Mean : 78.21	3500G : 3262	
	3rd Qu.: 6.000	3rd Qu.: 43.00	3rd Qu.:107.00	1000B : 3062	
	Max. :18.000	Max. :204.00	Max. :293.00	1500C : 2520	
				(Other): 3087	



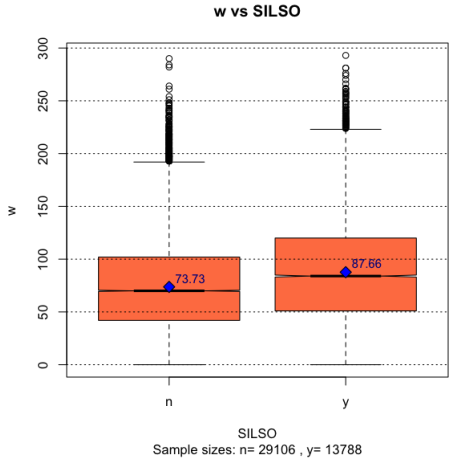
(a) Observed sunspot count histogram.



(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:

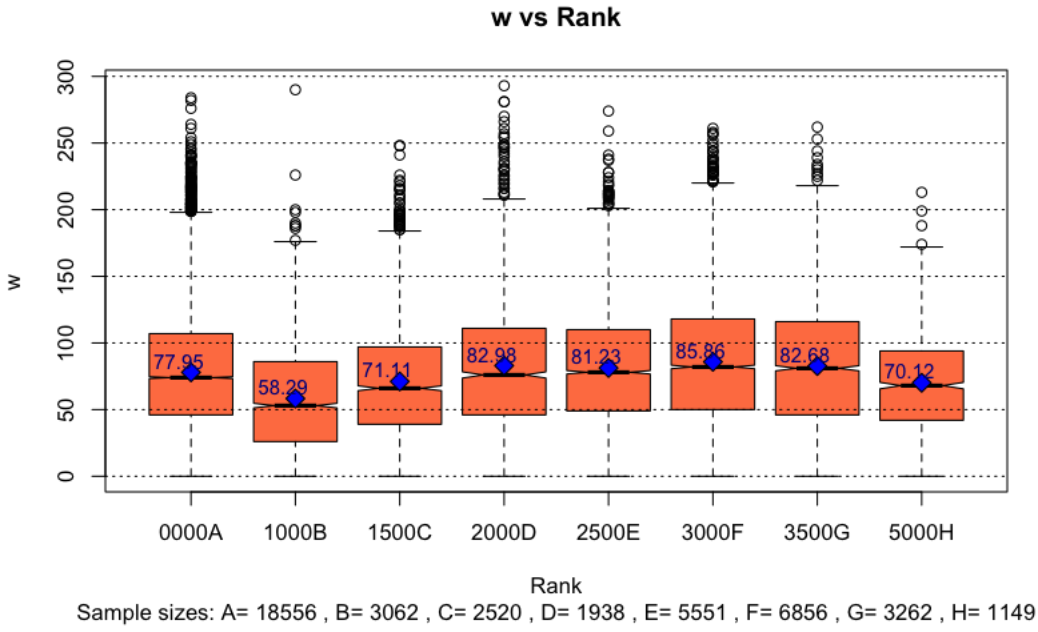


Figure 4: Box plot of sunspot count by rank.