

Monthly Report (00)

2015.07 Data Set

Wednesday 12th August, 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.

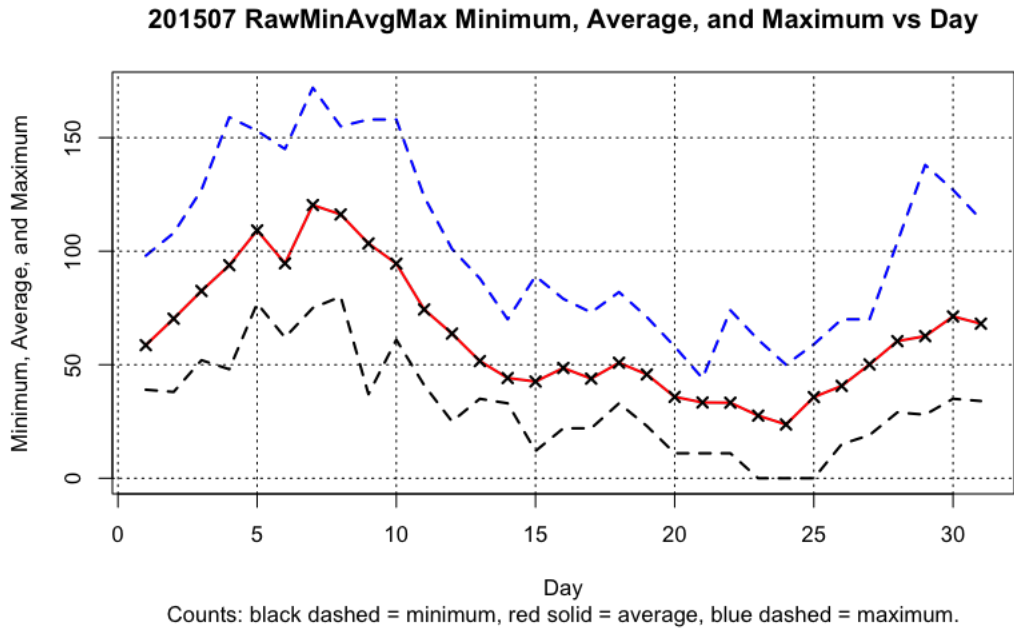


Figure 1: Raw average sunspot count by day of the month.

Table 1: 201507 Daily Raw Counts

Day	Submissions	Minimum	Average	Maximum
1.0000	34.0000	39.0000	58.6176	98.0000
2.0000	34.0000	38.0000	70.2812	108.0000
3.0000	40.0000	52.0000	82.5000	127.0000
4.0000	30.0000	48.0000	93.8276	159.0000
5.0000	33.0000	77.0000	109.1667	153.0000
6.0000	35.0000	62.0000	94.6471	145.0000
7.0000	30.0000	75.0000	120.3077	172.0000
8.0000	27.0000	80.0000	116.1923	155.0000
9.0000	29.0000	37.0000	103.4138	158.0000
10.0000	36.0000	61.0000	94.5143	158.0000
11.0000	44.0000	41.0000	74.3333	124.0000
12.0000	40.0000	25.0000	63.7059	101.0000
13.0000	34.0000	35.0000	51.6061	88.0000
14.0000	36.0000	33.0000	44.1429	70.0000
15.0000	32.0000	12.0000	42.6000	89.0000
16.0000	39.0000	22.0000	48.5405	79.0000
17.0000	31.0000	22.0000	43.8710	73.0000
18.0000	36.0000	33.0000	50.8571	82.0000
19.0000	34.0000	23.0000	45.7273	71.0000
20.0000	39.0000	11.0000	35.8378	58.0000
21.0000	33.0000	11.0000	33.3667	44.0000
22.0000	37.0000	11.0000	33.2571	74.0000
23.0000	39.0000	0.0000	27.5676	61.0000
24.0000	39.0000	0.0000	23.6842	50.0000
25.0000	39.0000	0.0000	35.7429	59.0000
26.0000	30.0000	15.0000	40.6667	70.0000
27.0000	38.0000	19.0000	50.1143	70.0000
28.0000	40.0000	29.0000	60.3846	104.0000
29.0000	37.0000	28.0000	62.5833	138.0000
30.0000	39.0000	35.0000	71.2571	127.0000
31.0000	33.0000	34.0000	68.0606	114.0000

3 Error Tables

Data are for the month of July 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 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
2010.05	22.5803	22.0695	23.0912
2010.06	18.1646	17.6833	18.6459
2010.07	19.9418	19.5053	20.3783
2010.08	19.6965	19.2215	20.1714
2010.09	23.5529	23.0551	24.0507
2010.10	23.1525	22.6574	23.6475
2010.11	25.2663	24.7005	25.8321
2010.12	23.7983	23.1215	24.4751
2011.01	71.9299	70.3241	73.5358
2011.02	60.9100	59.5401	62.2800
2011.03	69.2268	67.8078	70.6457
2011.04	75.3913	73.7802	77.0025
2011.05	77.3562	75.8089	78.9035
2011.06	65.6575	64.2804	67.0347

Continued on next page

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

ym	Ra	lci99	uci99
2011.07	70.2290	68.6892	71.7687
2011.08	72.5999	71.1813	74.0184
2011.09	84.1271	83.0266	85.2276
2011.10	82.3737	80.9714	83.7761
2011.11	88.4561	86.5954	90.3167
2011.12	80.8173	79.0693	82.5653
2012.01	74.1191	72.6636	75.5746
2012.02	60.7263	59.4516	62.0009
2012.03	71.8740	70.5980	73.1501
2012.04	75.6469	73.5272	77.7666
2012.05	81.4300	80.0118	82.8481
2012.06	68.9128	67.6935	70.1321
2012.07	74.6667	73.4171	75.9162
2012.08	73.5855	72.3630	74.8080
2012.09	85.2125	83.7653	86.6598
2012.10	84.9307	83.3322	86.5291
2012.11	92.3062	90.4669	94.1454
2012.12	82.0038	80.2731	83.7344
2013.01	83.8394	82.2672	85.4116
2013.02	70.0364	68.6704	71.4025
2013.03	78.8916	77.4081	80.3751
2013.04	87.9940	86.5178	89.4701
2013.05	89.2428	87.7126	90.7731
2013.06	76.0388	74.6914	77.3862
2013.07	80.2862	79.0371	81.5353
2013.08	81.2225	79.9571	82.4879
2013.09	93.2382	91.6354	94.8409
2013.10	91.5708	89.9556	93.1861
2013.11	99.5151	97.4578	101.5724
2013.12	90.6847	88.8626	92.5068
2014.01	99.6195	97.5323	101.7067
2014.02	83.5836	82.0058	85.1614
2014.03	97.1570	95.5195	98.7945
2014.04	106.9585	105.1519	108.7651
2014.05	108.0979	106.3770	109.8187
2014.06	92.2081	90.7153	93.7009
2014.07	98.7623	97.1605	100.3642
2014.08	99.2607	97.7726	100.7488

Continued on next page

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

ym	Ra	lci99	uci99
2014.09	115.0975	113.2724	116.9226
2014.10	112.5849	110.7093	114.4606
2014.11	123.0441	120.7115	125.3767
2014.12	109.5744	107.2287	111.9201
2015.01	69.7088	68.3653	71.0523
2015.02	58.3872	57.0402	59.7342
2015.03	66.7788	65.5624	67.9952
2015.04	74.3345	73.0231	75.6458
2015.05	74.8030	73.6085	75.9974
2015.06	64.2313	63.1393	65.3232
2015.07	67.3789	66.2358	68.5219

Loglinear Mixed Model Fit, AAVSO, and SILSO Values vs Sequence

Boxes and whiskers represent unprocessed counts

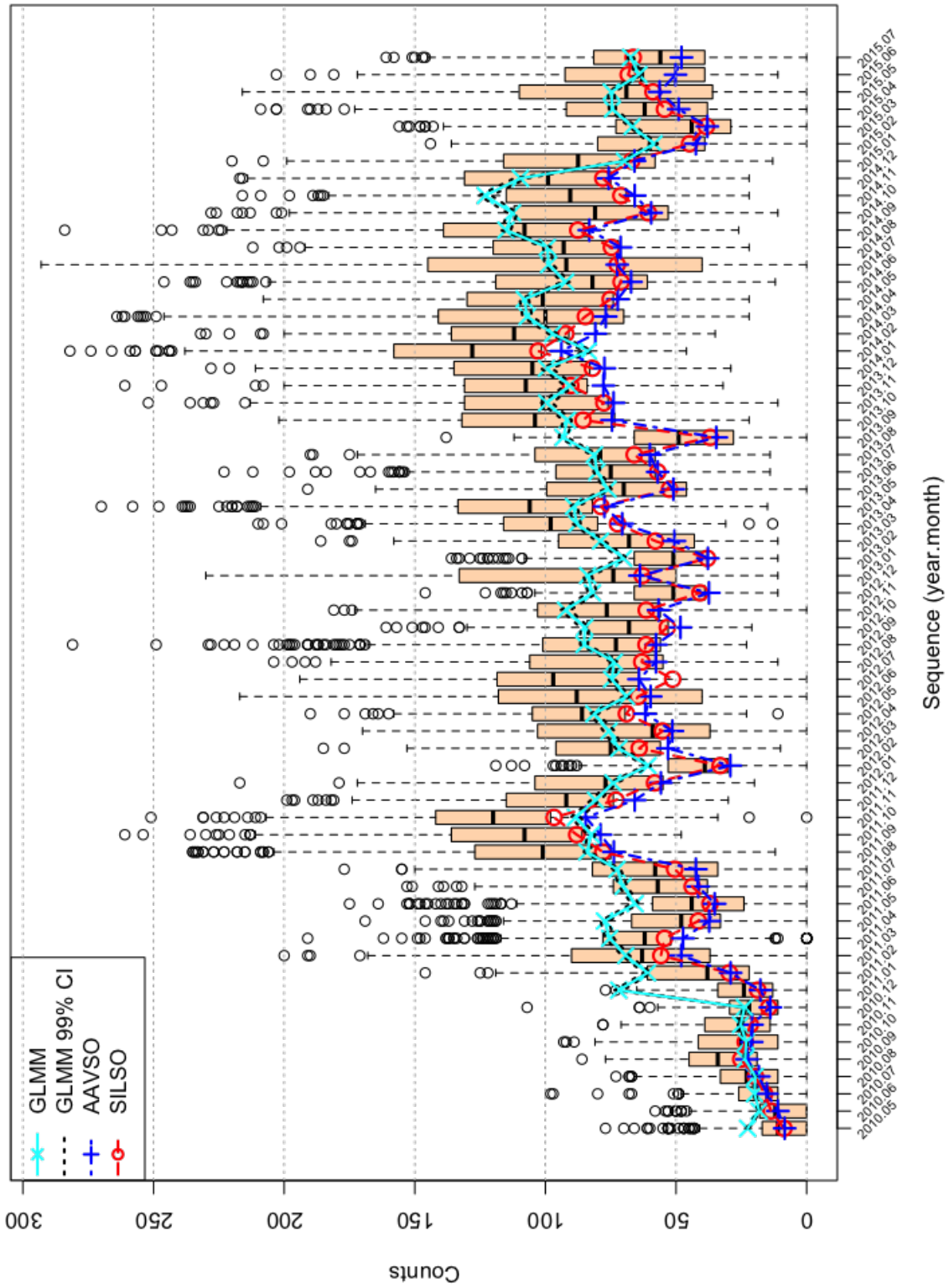


Figure 2: GLMM fitted data for R_a .

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: 201507 Parameter Estimates

	Estimate	Std. Error	t-value	$\Pr(> t)$
(Intercept)	3.1396	0.0446	70.3786	0.0000
seeF	-0.1827	0.0077	-23.7213	0.0000
seeG	-0.0937	0.0067	-13.9633	0.0000
seeP	-0.2929	0.0113	-25.8655	0.0000
r1000B	-0.0510	0.0831	-0.6136	0.5395
r1500C	0.0431	0.1269	0.3401	0.7338
r2000D	0.0849	0.1547	0.5490	0.5830
r2500E	0.0017	0.1051	0.0166	0.9868
r3000F	0.0628	0.1023	0.6134	0.5396
r3500G	0.1222	0.1531	0.7985	0.4246
r5000H	-0.1039	0.2118	-0.4907	0.6236
silsoy	0.1230	0.0738	1.6676	0.0954
year2011	1.2251	0.0154	79.6726	0.0000
year2012	1.2440	0.0153	81.1492	0.0000
year2013	1.3417	0.0153	87.7602	0.0000
year2014	1.5317	0.0152	100.8345	0.0000
year2015	1.1729	0.0165	70.9529	0.0000
mon2	-0.1867	0.0128	-14.5459	0.0000
mon3	-0.0531	0.0116	-4.5596	0.0000
mon4	0.0459	0.0117	3.9239	0.0001
mon5	0.0591	0.0111	5.3349	0.0000
mon6	-0.1095	0.0116	-9.4283	0.0000
mon7	-0.0461	0.0112	-4.1119	0.0000
mon8	-0.0292	0.0115	-2.5398	0.0111
mon9	0.1189	0.0110	10.8600	0.0000
mon10	0.1040	0.0116	8.9761	0.0000
mon11	0.1968	0.0121	16.3144	0.0000
mon12	0.0918	0.0125	7.3134	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.

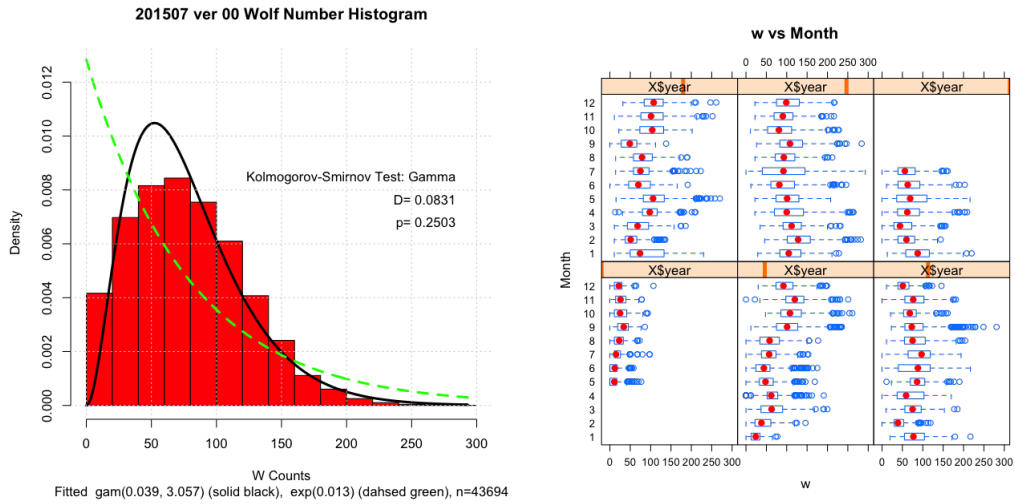
5 Supporting Information

Table 4: 201507 Summary of Sunspot Numbers

obs	jd	year	mon	day
ARAG : 1854	Min. :1721096	Min. :2010	Min. : 1.000	Min. : 1.00
CHAG : 1676	1st Qu.:2455856	1st Qu.:2011	1st Qu.: 4.000	1st Qu.: 8.00
BRAB : 1646	Median :2456326	Median :2013	Median : 7.000	Median :16.00
BROB : 1474	Mean :2455912	Mean :2013	Mean : 6.542	Mean :15.74
DUBF : 1393	3rd Qu.:2456776	3rd Qu.:2014	3rd Qu.: 9.000	3rd Qu.:23.00
HOWR : 1349	Max. :2457235	Max. :2015	Max. :12.000	Max. :31.00
(Other):34302				

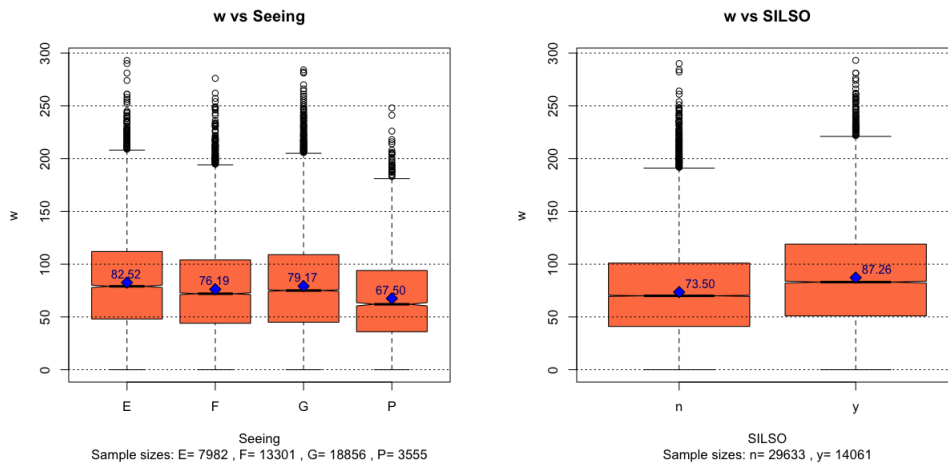
Table 5: Summary of Sunspot Numbers

see	g	s	w	r	silso
E: 7982	Min. : 0.000	Min. : 0.00	Min. : 0.00	0000A :18930	n:29633
F:13301	1st Qu.: 3.000	1st Qu.: 12.00	1st Qu.: 44.00	3000F : 6963	y:14061
G:18856	Median : 5.000	Median : 25.00	Median : 74.00	2500E : 5645	
P: 3555	Mean : 4.786	Mean : 30.07	Mean : 77.93	3500G : 3322	
	3rd Qu.: 6.000	3rd Qu.: 43.00	3rd Qu.:107.00	1000B : 3152	
	Max. :18.000	Max. :204.00	Max. :293.00	1500C : 2574	
				(Other): 3108	



(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:

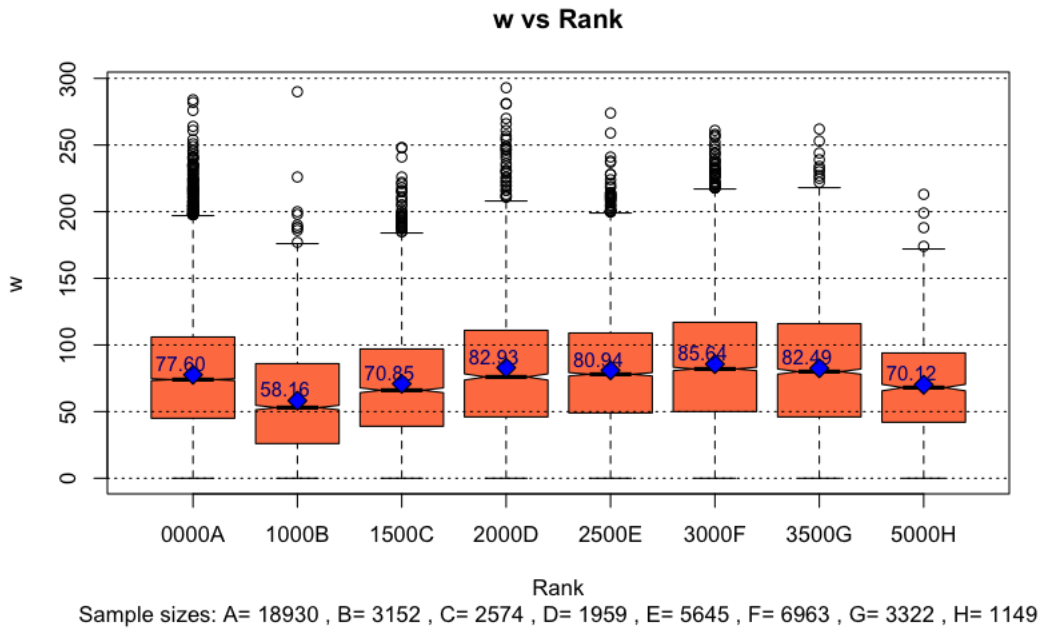


Figure 4: Box plot of sunspot count by rank.