

Monthly Report (00)

2015.10 Data Set

Monday 16th November, 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.

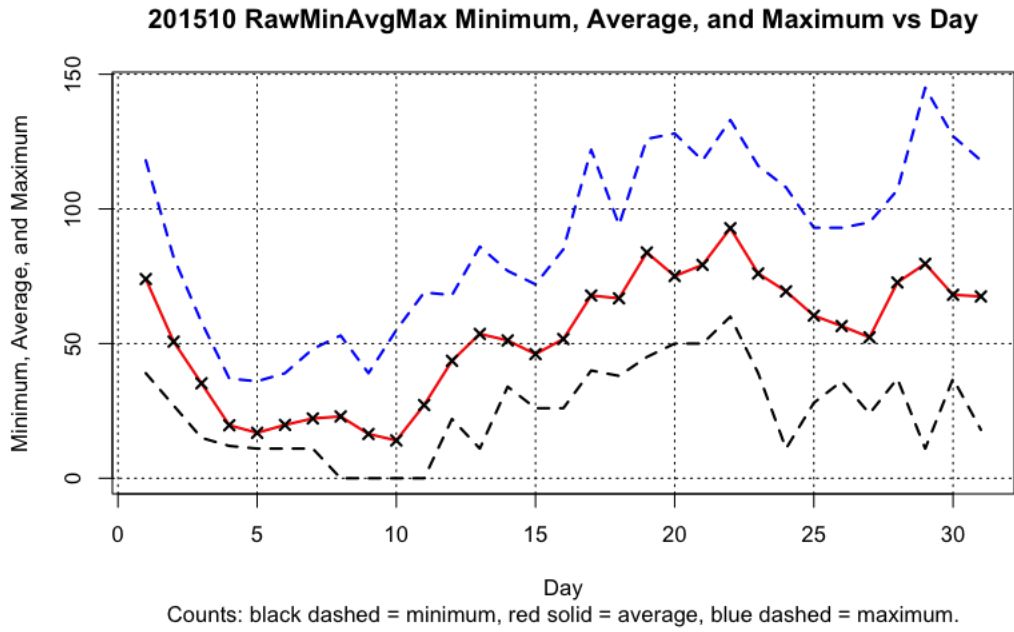


Figure 1: Raw average sunspot count by day of the month.

Table 1: 201510 Daily Raw Counts

Day	Submissions	Minimum	Average	Maximum
1.0000	27.0000	39.0000	73.9583	118.0000
2.0000	24.0000	27.0000	50.7273	82.0000
3.0000	22.0000	15.0000	35.2500	58.0000
4.0000	31.0000	12.0000	19.6786	37.0000
5.0000	33.0000	11.0000	16.8667	36.0000
6.0000	33.0000	11.0000	19.8000	39.0000
7.0000	40.0000	11.0000	22.1944	48.0000
8.0000	40.0000	0.0000	22.9429	53.0000
9.0000	40.0000	0.0000	16.4286	39.0000
10.0000	37.0000	0.0000	14.0323	55.0000
11.0000	37.0000	0.0000	27.1875	69.0000
12.0000	37.0000	22.0000	43.6061	68.0000
13.0000	32.0000	11.0000	53.5862	86.0000
14.0000	34.0000	34.0000	51.1613	77.0000
15.0000	35.0000	26.0000	46.0938	72.0000
16.0000	32.0000	26.0000	51.6897	85.0000
17.0000	27.0000	40.0000	67.8333	122.0000
18.0000	36.0000	38.0000	66.8065	94.0000
19.0000	31.0000	45.0000	83.8276	126.0000
20.0000	29.0000	50.0000	75.0741	128.0000
21.0000	27.0000	50.0000	79.2083	118.0000
22.0000	30.0000	60.0000	92.8214	133.0000
23.0000	28.0000	39.0000	76.0385	116.0000
24.0000	18.0000	11.0000	69.4000	108.0000
25.0000	28.0000	28.0000	60.3750	93.0000
26.0000	35.0000	36.0000	56.5152	93.0000
27.0000	33.0000	24.0000	52.3226	95.0000
28.0000	25.0000	37.0000	72.6818	107.0000
29.0000	32.0000	11.0000	79.6129	145.0000
30.0000	32.0000	37.0000	68.1071	127.0000
31.0000	32.0000	18.0000	67.5000	118.0000

3 Error Tables

Data are for the month of October 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	23.0011	22.4799	23.5223
2010.06	18.5176	18.0273	19.0079
2010.07	20.3094	19.8646	20.7542
2010.08	19.3874	18.9189	19.8560
2010.09	23.3592	22.8636	23.8549
2010.10	22.3697	21.8902	22.8492
2010.11	25.4085	24.8376	25.9794
2010.12	23.9429	23.2618	24.6240
2011.01	73.1047	71.4724	74.7370
2011.02	61.7504	60.3606	63.1403
2011.03	70.3110	68.8680	71.7540
2011.04	76.8694	75.2268	78.5120
2011.05	78.5713	76.9990	80.1437
2011.06	66.6789	65.2803	68.0775

Continued on next page

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

ym	Ra	lci99	uci99
2011.07	71.2850	69.7241	72.8459
2011.08	71.2893	69.8984	72.6802
2011.09	83.2382	82.1530	84.3234
2011.10	79.3150	77.9679	80.6621
2011.11	88.5447	86.6776	90.4119
2011.12	80.9083	79.1574	82.6592
2012.01	75.1913	73.7132	76.6693
2012.02	61.4336	60.1418	62.7253
2012.03	72.8509	71.5575	74.1444
2012.04	76.7473	74.5570	78.9377
2012.05	82.6070	81.1716	84.0423
2012.06	69.8507	68.6155	71.0860
2012.07	75.6730	74.4062	76.9399
2012.08	72.1234	70.9244	73.3223
2012.09	84.1083	82.6788	85.5377
2012.10	81.5934	80.0611	83.1258
2012.11	92.3064	90.4691	94.1436
2012.12	81.9971	80.2718	83.7223
2013.01	84.9356	83.3473	86.5238
2013.02	70.8557	69.4798	72.2317
2013.03	79.9375	78.4400	81.4351
2013.04	89.4004	87.8986	90.9023
2013.05	90.3273	88.7753	91.8793
2013.06	76.9951	75.6288	78.3615
2013.07	81.2863	80.0196	82.5530
2013.08	79.5307	78.2907	80.7708
2013.09	91.9609	90.3793	93.5425
2013.10	87.7892	86.2408	89.3377
2013.11	99.3822	97.3271	101.4374
2013.12	90.6186	88.8025	92.4348
2014.01	100.9685	98.8549	103.0820
2014.02	84.4747	82.8789	86.0704
2014.03	98.4756	96.8234	100.1278
2014.04	108.7241	106.8846	110.5637
2014.05	109.4898	107.7486	111.2309
2014.06	93.3831	91.8716	94.8947
2014.07	99.9893	98.3684	101.6101
2014.08	97.2474	95.7880	98.7068

Continued on next page

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

ym	Ra	lci99	uci99
2014.09	113.4588	111.6532	115.2644
2014.10	108.0534	106.2553	109.8514
2014.11	122.9568	120.6320	125.2817
2014.12	109.4509	107.1130	111.7889
2015.01	64.8877	63.6441	66.1313
2015.02	54.2121	52.9690	55.4553
2015.03	62.1342	61.0074	63.2610
2015.04	69.4477	68.2288	70.6667
2015.05	69.6292	68.5139	70.7445
2015.06	59.7802	58.7710	60.7894
2015.07	62.6163	61.5547	63.6779
2015.08	62.3712	61.3550	63.3873
2015.09	72.5925	71.4185	73.7665
2015.10	69.0841	67.9106	70.2576

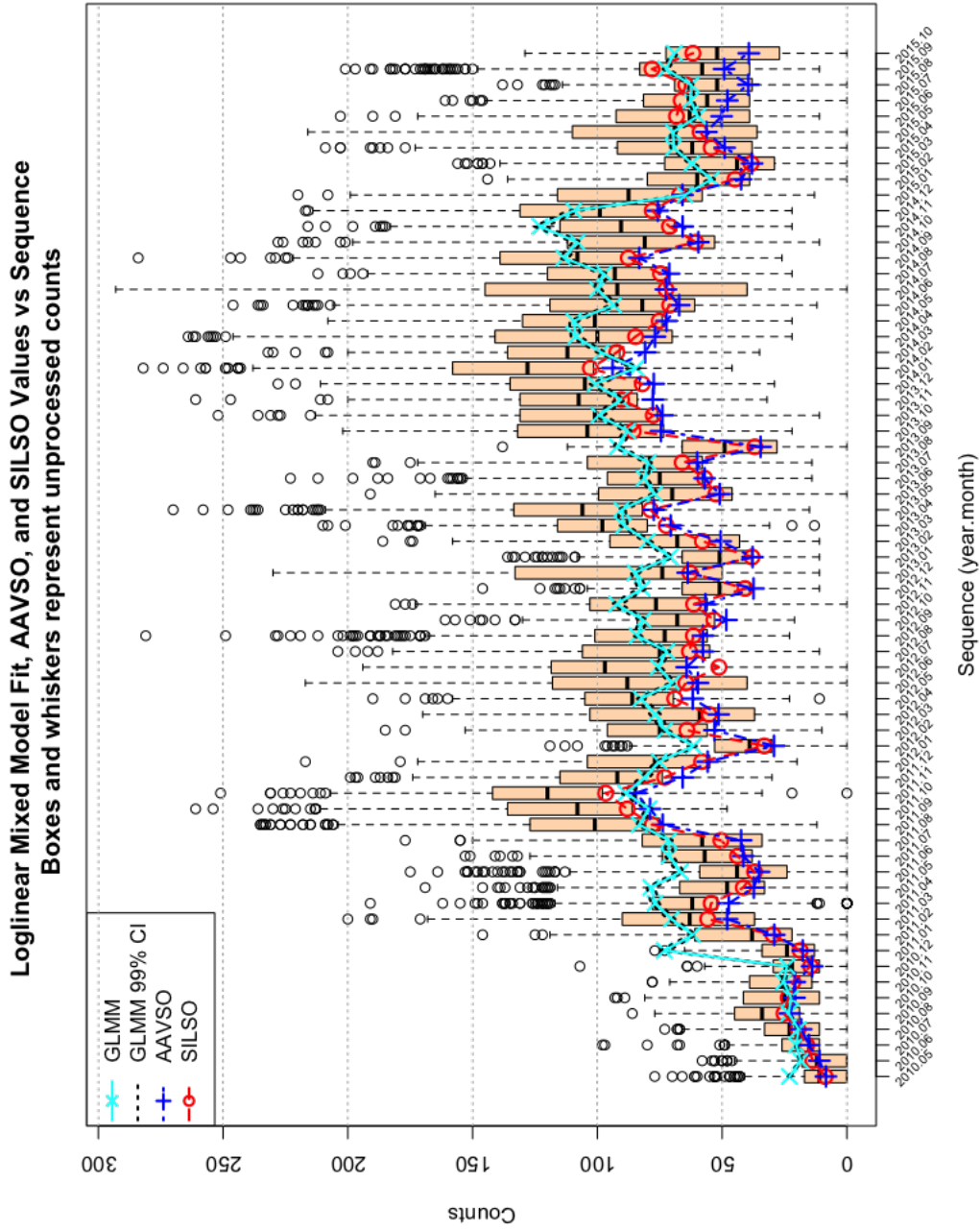


Figure 2: GLMM fitted data for R_a .

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: 201510 Parameter Estimates

	Estimate	Std. Error	t-value	$\Pr(> t)$
(Intercept)	3.1654	0.0445	71.1809	0.0000
seeF	-0.1827	0.0075	-24.2208	0.0000
seeG	-0.0951	0.0066	-14.4947	0.0000
seeP	-0.2933	0.0111	-26.3961	0.0000
r1000B	-0.0602	0.0828	-0.7265	0.4675
r1500C	0.0319	0.1265	0.2525	0.8007
r2000D	0.0846	0.1542	0.5486	0.5833
r2500E	-0.0004	0.1048	-0.0034	0.9973
r3000F	0.0606	0.1020	0.5942	0.5524
r3500G	0.1188	0.1526	0.7780	0.4366
r5000H	-0.1133	0.2111	-0.5365	0.5916
silsoy	0.1191	0.0735	1.6188	0.1055
year2011	1.2204	0.0154	79.4980	0.0000
year2012	1.2377	0.0153	80.8763	0.0000
year2013	1.3344	0.0153	87.4378	0.0000
year2014	1.5244	0.0152	100.5386	0.0000
year2015	1.0797	0.0158	68.2321	0.0000
mon2	-0.1889	0.0128	-14.7309	0.0000
mon3	-0.0532	0.0116	-4.5754	0.0000
mon4	0.0497	0.0117	4.2479	0.0000
mon5	0.0590	0.0111	5.3362	0.0000
mon6	-0.1095	0.0116	-9.4437	0.0000
mon7	-0.0465	0.0112	-4.1582	0.0000
mon8	-0.0633	0.0111	-5.7155	0.0000
mon9	0.0923	0.0106	8.6856	0.0000
mon10	0.0496	0.0112	4.4282	0.0000
mon11	0.1829	0.0120	15.2201	0.0000
mon12	0.0772	0.0125	6.1757	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.

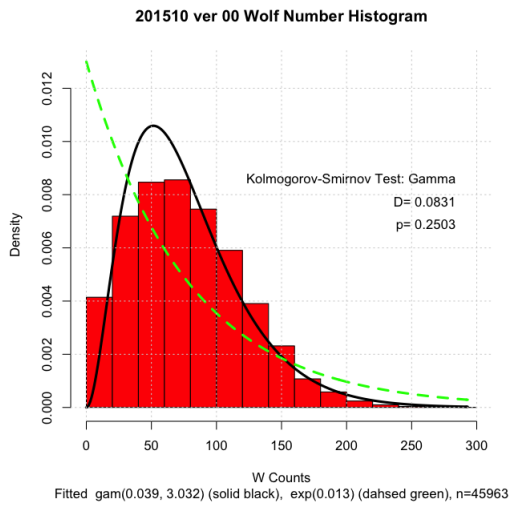
5 Supporting Information

Table 4: 201510 Summary of Sunspot Numbers

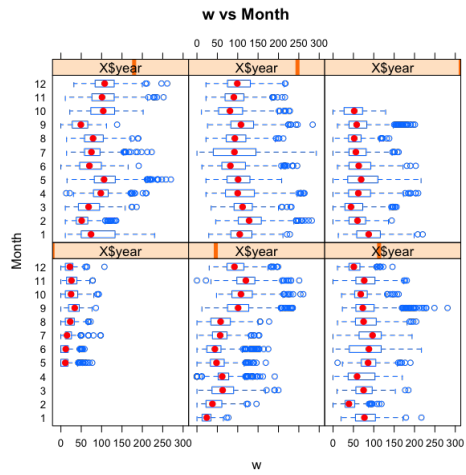
obs	jd	year	mon	day
ARAG : 1944	Min. :1721096	Min. :2010	Min. : 1.000	Min. : 1.00
CHAG : 1764	1st Qu.:2455883	1st Qu.:2011	1st Qu.: 4.000	1st Qu.: 8.00
BRAB : 1734	Median :2456379	Median :2013	Median : 7.000	Median :16.00
BROB : 1561	Mean :2455980	Mean :2013	Mean : 6.661	Mean :15.73
DUBF : 1469	3rd Qu.:2456838	3rd Qu.:2014	3rd Qu.: 9.000	3rd Qu.:23.00
HOWR : 1426	Max. :2457327	Max. :2015	Max. :12.000	Max. :31.00
(Other):36065				

Table 5: Summary of Sunspot Numbers

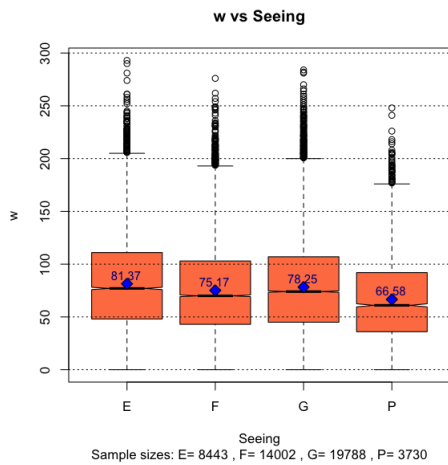
see	g	s	w	r	silso
E: 8443	Min. : 0.000	Min. : 0.0	Min. : 0.00	0000A :19849	n:31179
F:14002	1st Qu.: 3.000	1st Qu.: 12.0	1st Qu.: 44.00	3000F : 7352	y:14784
G:19788	Median : 5.000	Median : 24.0	Median : 72.00	2500E : 5932	
P: 3730	Mean : 4.714	Mean : 29.8	Mean : 76.94	3500G : 3498	
	3rd Qu.: 6.000	3rd Qu.: 43.0	3rd Qu.:106.00	1000B : 3354	
	Max. :18.000	Max. :204.0	Max. :293.00	1500C : 2747	
				(Other): 3231	



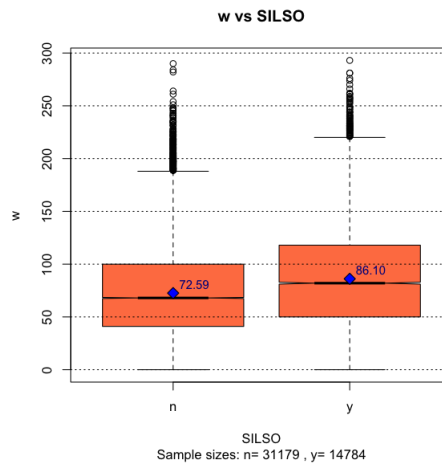
(a) Observed sunspot count histogram.



(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:

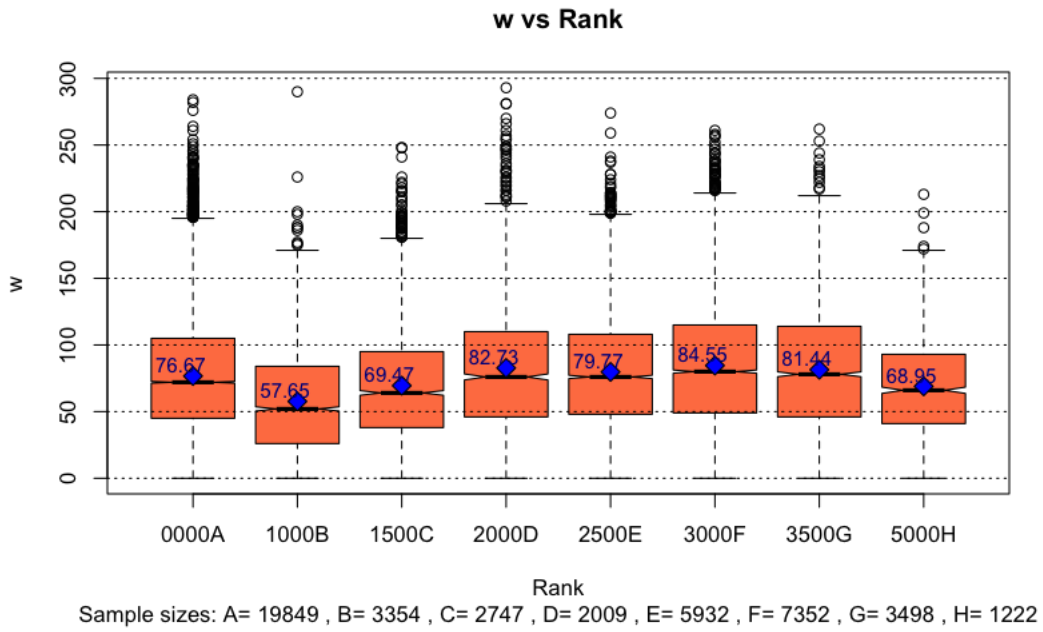


Figure 4: Box plot of sunspot count by rank.