Monthly Report (00)

201707 Data Set

Monday 14th August, 2017

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 (GLMM05) 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.

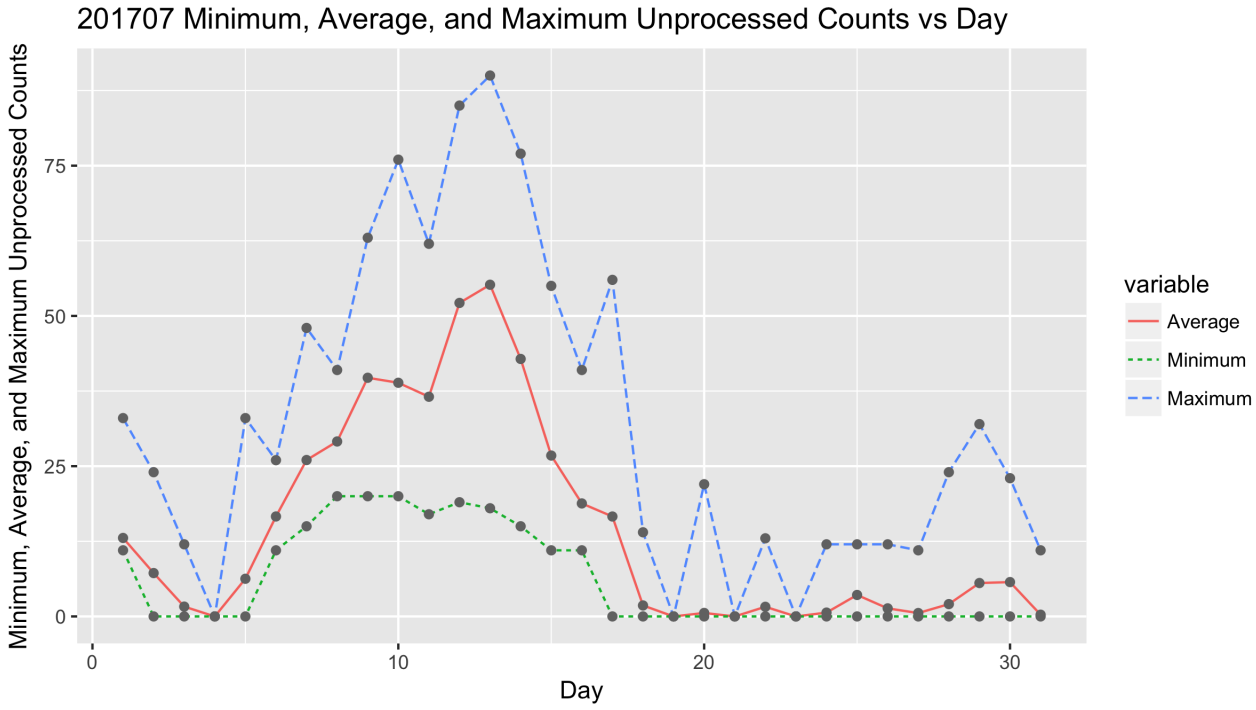

Figure 1: Raw average sunspot count by day of the month.

Table 1: 201707 Daily Raw Counts

Day	Submissions	Minimum	Average	Maximum
1.0000	34.0000	11.0000	13.0588	33.0000
2.0000	42.0000	0.0000	7.2143	24.0000
3.0000	41.0000	0.0000	1.6341	12.0000
4.0000	43.0000	0.0000	0.0000	0.0000
5.0000	41.0000	0.0000	6.2683	33.0000
6.0000	38.0000	11.0000	16.6316	26.0000
7.0000	40.0000	15.0000	26.0250	48.0000
8.0000	45.0000	20.0000	29.1111	41.0000
9.0000	45.0000	20.0000	39.7111	63.0000
10.0000	41.0000	20.0000	38.8780	76.0000
11.0000	38.0000	17.0000	36.5526	62.0000
12.0000	37.0000	19.0000	52.1622	85.0000
13.0000	37.0000	18.0000	55.1892	90.0000
14.0000	38.0000	15.0000	42.8421	77.0000
15.0000	43.0000	11.0000	26.7674	55.0000
16.0000	40.0000	11.0000	18.8000	41.0000
17.0000	40.0000	0.0000	16.6250	56.0000
18.0000	40.0000	0.0000	1.8250	14.0000
19.0000	38.0000	0.0000	0.0000	0.0000
20.0000	38.0000	0.0000	0.5789	22.0000
21.0000	41.0000	0.0000	0.0000	0.0000
22.0000	43.0000	0.0000	1.6047	13.0000
23.0000	36.0000	0.0000	0.0000	0.0000
24.0000	36.0000	0.0000	0.6389	12.0000
25.0000	35.0000	0.0000	3.5714	12.0000
26.0000	35.0000	0.0000	1.3429	12.0000
27.0000	39.0000	0.0000	0.5641	11.0000
28.0000	40.0000	0.0000	2.0500	24.0000
29.0000	34.0000	0.0000	5.5588	32.0000
30.0000	35.0000	0.0000	5.7143	23.0000
31.0000	38.0000	0.0000	0.2895	11.0000

3 Error Tables

Data are for the month of July 2017. 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	aavso	silso
2010.05	23.7466	23.2033	24.2899	8.4000	8.7000
2010.06	18.4811	17.9881	18.9740	11.0000	13.6000
2010.07	20.3990	19.9479	20.8501	15.2000	16.1000
2010.08	20.3350	19.8397	20.8303	18.3000	19.6000
2010.09	23.9343	23.4204	24.4481	22.8000	25.2000
2010.10	22.7555	22.2652	23.2458	21.0000	23.5000
2010.11	23.4741	22.9451	24.0031	20.9000	21.6000
2010.12	22.4513	21.8039	23.0987	13.9000	14.5000
2011.01	76.3116	74.5935	78.0298	17.7000	18.7000
2011.02	66.5959	65.0952	68.0965	29.1000	29.6000
2011.03	71.4319	69.9554	72.9084	48.0000	55.8000
2011.04	78.7743	77.0908	80.4579	47.3000	54.4000
2011.05	80.0353	78.4389	81.6316	37.3000	41.5000
2011.06	65.6095	64.2292	66.9898	35.2000	37.0000
2011.07	70.5675	69.0141	72.1210	41.5000	43.8000

Continued on next page

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

ym	Ra	lci99	uci99	aavso	silso
2011.08	73.7287	72.2827	75.1748	42.4000	50.5000
2011.09	83.8806	82.7771	84.9841	73.8000	78.0000
2011.10	79.1046	77.7459	80.4633	78.9000	88.0000
2011.11	80.3240	78.6003	82.0477	84.6000	96.7000
2011.12	74.6878	73.0410	76.3346	65.8000	73.0000
2012.01	78.4626	76.9039	80.0213	55.8000	58.2000
2012.02	66.2561	64.8457	67.6666	29.2000	33.1000
2012.03	73.9096	72.5875	75.2317	53.1000	64.1000
2012.04	78.5376	76.2185	80.8568	51.4000	55.2000
2012.05	83.9830	82.5166	85.4493	61.8000	69.0000
2012.06	68.6634	67.4479	69.8789	59.7000	64.5000
2012.07	74.8161	73.5560	76.0762	64.2000	51.3000
2012.08	74.4551	73.2097	75.7005	57.7000	63.1000
2012.09	84.8100	83.3576	86.2625	57.7000	61.5000
2012.10	81.5644	80.0219	83.1069	48.3000	53.3000
2012.11	83.8843	82.2016	85.5671	56.7000	61.4000
2012.12	75.7535	74.1588	77.3482	37.4000	40.8000
2013.01	88.3525	86.7008	90.0042	63.8000	62.9000
2013.02	76.3830	74.8931	77.8729	37.8000	38.0000
2013.03	81.1442	79.6161	82.6722	50.6000	57.9000
2013.04	91.4051	89.8653	92.9448	70.6000	72.4000
2013.05	91.6178	90.0381	93.1976	77.4000	78.7000
2013.06	75.5381	74.1961	76.8801	51.0000	52.5000
2013.07	80.3053	79.0543	81.5562	57.0000	57.0000
2013.08	82.0439	80.7614	83.3263	60.0000	66.0000
2013.09	92.5562	90.9546	94.1577	34.6000	36.9000
2013.10	87.4374	85.8850	88.9899	74.5000	85.6000
2013.11	90.0560	88.1828	91.9292	73.9000	77.6000
2013.12	83.3862	81.7086	85.0637	77.8000	90.3000
2014.01	104.9341	102.7257	107.1424	77.4000	82.0000
2014.02	90.6650	88.9409	92.3891	93.9000	102.8000
2014.03	99.7871	98.1012	101.4730	80.9000	92.2000
2014.04	111.1004	109.2127	112.9882	76.9000	84.7000
2014.05	110.8722	109.0948	112.6496	72.3000	75.2000
2014.06	91.4421	89.9584	92.9259	67.2000	71.0000
2014.07	98.5290	96.9206	100.1374	72.5000	72.5000
2014.08	100.1624	98.6571	101.6677	71.2000	74.7000
2014.09	114.1794	112.3356	116.0232	83.2000	87.6000
2014.10	107.6722	105.8702	109.4742	59.5000	60.6000

Continued on next page

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

ym	Ra	lci99	uci99	aavso	silso
2014.11	111.4147	109.3139	113.5156	65.8000	71.1000
2014.12	100.8536	98.6636	103.0435	75.8000	78.0000
2015.01	64.0296	62.7951	65.2641	65.9000	67.0000
2015.02	55.3050	54.0131	56.5969	42.4000	44.8000
2015.03	59.6552	58.5591	60.7514	38.0000	38.4000
2015.04	67.2293	66.0361	68.4224	49.0000	54.4000
2015.05	66.7170	65.6444	67.7895	56.3000	58.8000
2015.06	55.4954	54.5485	56.4423	50.2000	68.3000
2015.07	58.6206	57.6180	59.6232	47.9000	65.8000
2015.08	61.0591	60.0579	62.0603	39.5000	57.2000
2015.09	69.3071	68.1770	70.4372	49.2000	72.1000
2015.10	65.3228	64.2083	66.4372	39.3000	48.3000
2015.11	68.0992	67.1766	69.0217	39.6000	55.9000
2015.12	61.1723	59.9398	62.4047	36.4000	44.8000
2016.01	35.7203	35.0852	36.3554	33.7000	43.3000
2016.02	30.2190	29.6194	30.8186	38.3000	46.8000
2016.03	32.4787	31.8808	33.0766	30.5000	38.9000
2016.04	35.9783	35.3386	36.6181	26.6000	30.9000
2016.05	36.7004	36.0724	37.3284	33.7000	48.4000
2016.06	30.1403	29.6616	30.6190	13.1000	19.5000
2016.07	32.5105	32.0077	33.0134	21.2000	27.5000
2016.08	33.5225	32.9698	34.0752	33.0000	47.9000
2016.09	37.8028	37.1747	38.4309	27.7000	37.1000
2016.10	35.8741	35.2521	36.4962	22.7000	31.7000
2016.11	36.9826	36.3449	37.6203	14.0000	22.2000
2016.12	33.5249	32.8518	34.1980	11.1000	20.0000
2017.01	20.0428	19.6611	20.4246	18.4000	26.2000
2017.02	16.7981	16.4686	17.1277	14.4000	20.6000
2017.03	18.4617	18.1347	18.7886	11.3000	15.5000
2017.04	20.8386	20.4983	21.1789	21.6000	33.2000
2017.05	20.7188	20.3805	21.0570	12.5000	18.1000
2017.06	16.7071	16.4687	16.9456	15.5000	19.3000
2017.07	18.1638	17.9184	18.4091	11.5000	16.3000

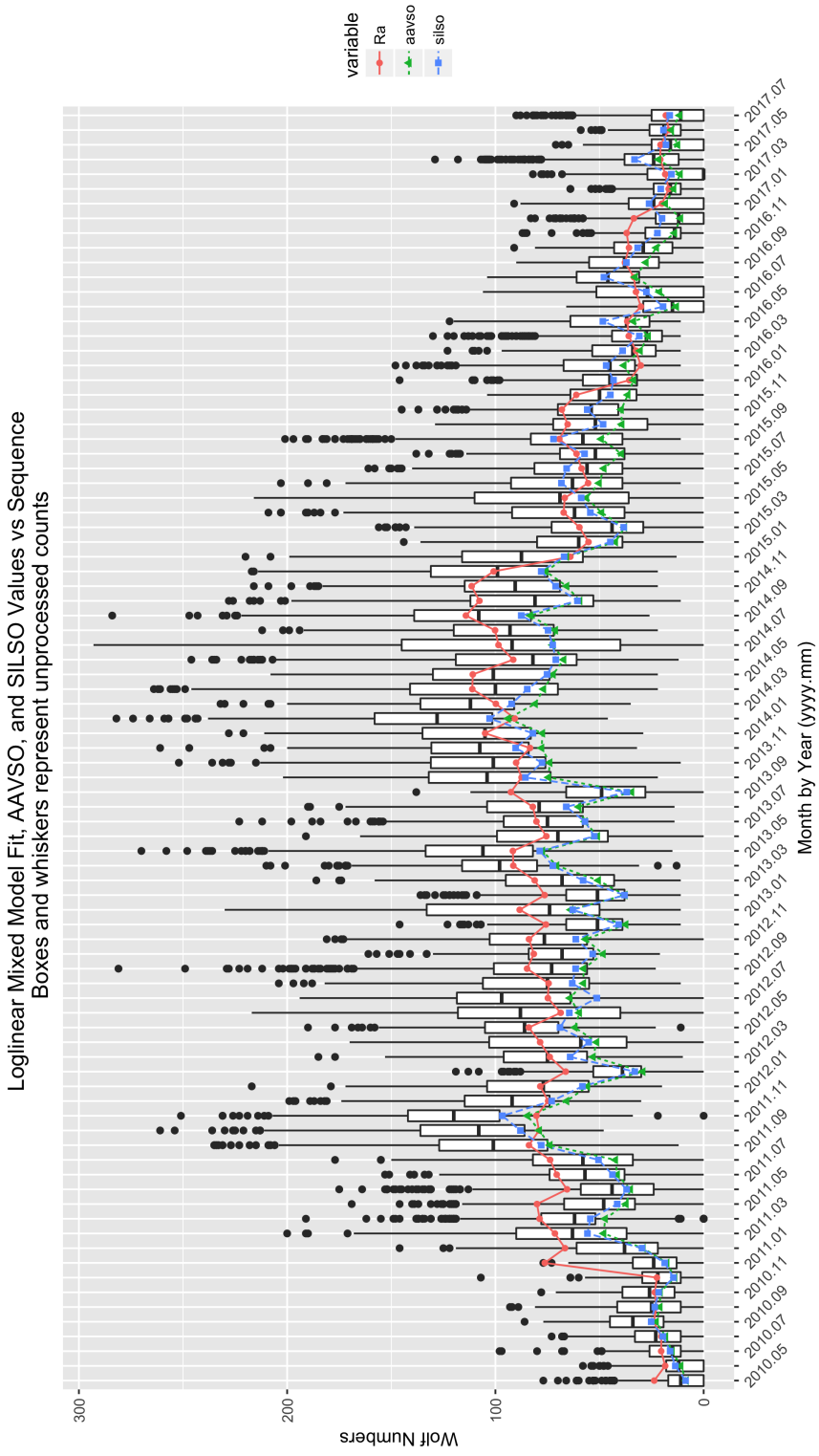

Figure 2: GLMM fitted data for R_a. AAVSO data: https://www.aavso.org/category/tags/solar-bulletin. SILSO data: WDC-SILSO, Royal Observatory of Belgium, Brussels

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: 201707 Parameter Estimates

	Estimate	Std. Error	t-value	Pr(>|t|)
(Intercept)	3.2230	0.0328	98.1441	0.0000
seeF	-0.1885	0.0072	-26.3095	0.0000
seeG	-0.1013	0.0062	-16.2712	0.0000
seeP	-0.2921	0.0105	-27.7923	0.0000
silso1	0.0623	0.0428	1.4552	0.1456
year2011	1.2040	0.0154	77.9534	0.0000
year2012	1.2225	0.0154	79.4161	0.0000
year2013	1.3172	0.0153	85.8153	0.0000
year2014	1.5059	0.0152	98.7614	0.0000
year2015	1.0077	0.0156	64.4153	0.0000
year2016	0.4026	0.0166	24.1984	0.0000
year2017	-0.1739	0.0200	-8.6844	0.0000
mon2	-0.1556	0.0119	-13.1100	0.0000
mon3	-0.0789	0.0109	-7.2149	0.0000
mon4	0.0315	0.0109	2.8829	0.0039
mon5	0.0335	0.0104	3.2246	0.0013
mon6	-0.1688	0.0109	-15.4370	0.0000
mon7	-0.0992	0.0105	-9.4121	0.0000
mon8	-0.0719	0.0105	-6.8566	0.0000
mon9	0.0590	0.0101	5.8358	0.0000
mon10	0.0062	0.0107	0.5822	0.5604
mon11	0.0443	0.0109	4.0532	0.0001
mon12	-0.0450	0.0116	-3.8718	0.0001

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: 201707 Summary of Sunspot Numbers

obs	jd	year	mon	day
ARAG : 2584	Min. :1721096	Min. :2010	Min. : 1.000	Min. : 1.00
CHAG : 2371	1st Qu.:2456070	1st Qu.:2012	1st Qu.: 4.000	1st Qu.: 8.00
BRAB : 2365	Median :2456677	Median :2014	Median : 7.000	Median :16.00
BROB : 2096	Mean :2456396	Mean :2014	Mean : 6.578	Mean :15.72
KNJS : 1981	3rd Qu.:2457319	3rd Qu.:2015	3rd Qu.: 9.000	3rd Qu.:23.00
HOWR : 1967	Max. :2457966	Max. :2017	Max. :12.000	Max. :31.00
(Other):47692				

Table 5: Summary of Sunspot Numbers

see	g	s	w	r	silso
E:11249	Min. : 0.000	Min. : 0.0	Min. : 0.00	0000A :24894	Min. :0.0000
F:18726	1st Qu.: 2.000	1st Qu.: 8.0	1st Qu.: 30.00	3000F : 9764	1st Qu.:0.0000
G:26099	Median : 4.000	Median : 19.0	Median : 59.00	2500E : 7766	Median :0.0000
P: 4982	Mean : 4.038	Mean : 25.1	Mean : 65.48	3500G : 4618	Mean :0.3284
	3rd Qu.: 6.000	3rd Qu.: 36.0	3rd Qu.: 94.00	1000B : 4228	3rd Qu.:1.0000
	Max. :18.000	Max. :204.0	Max. :293.00	1500C : 3059	Max. :1.0000
				(Other): 6727	

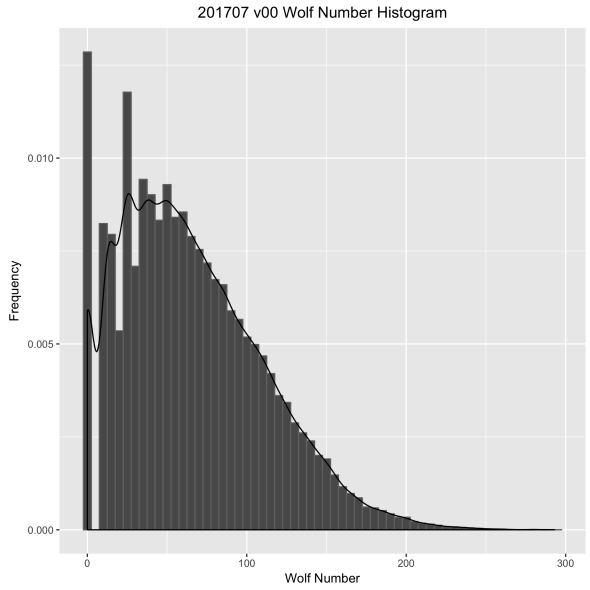

Figure 3: Box plots of raw Wolf number (w) by observer rank.

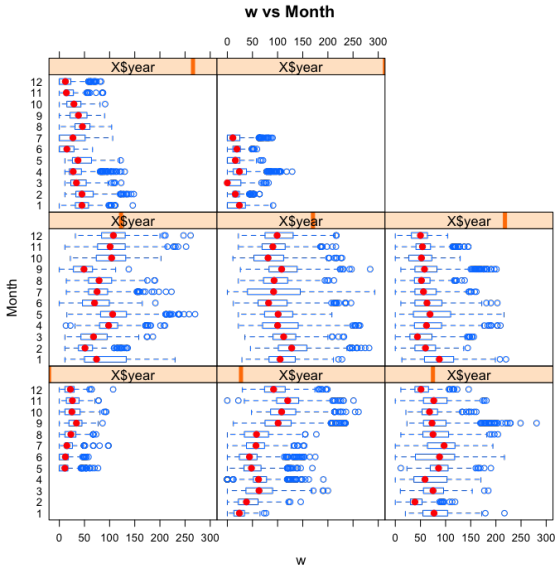

Figure 4: Box plots of raw Wolf number (w) by month and year.

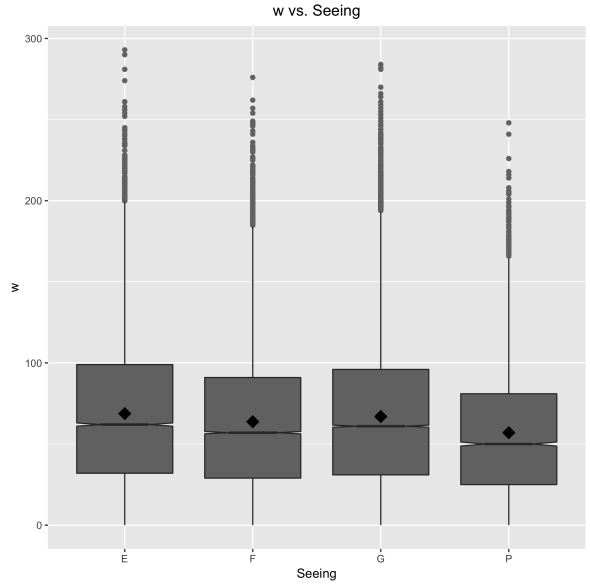

Figure 5: Box plots of raw Wolf number (w) by seeing condition.

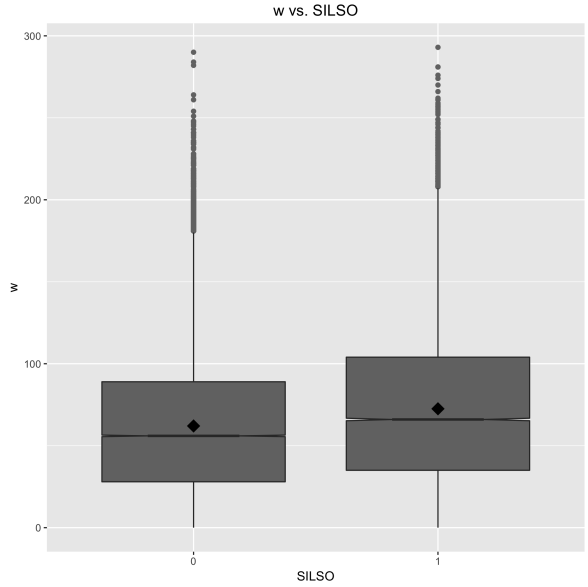

Figure 6: Box plots of raw Wolf number (w) by organization.

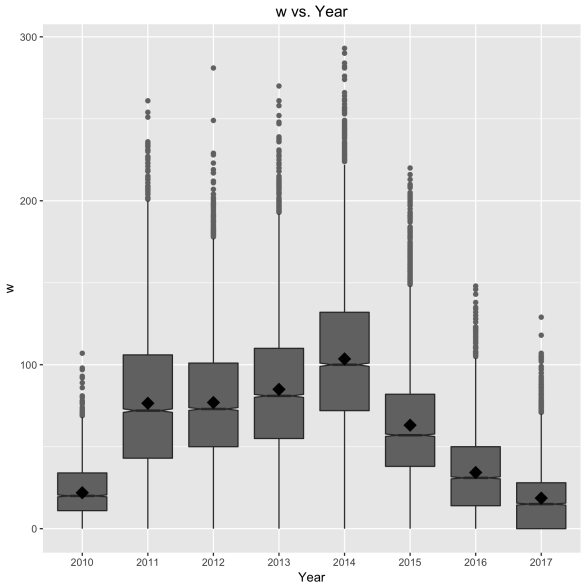

Figure 7: Box plots of raw Wolf number (w) by year.