Monthly Report (00)

2017.01 Data Set

Wednesday 15th February, 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 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.

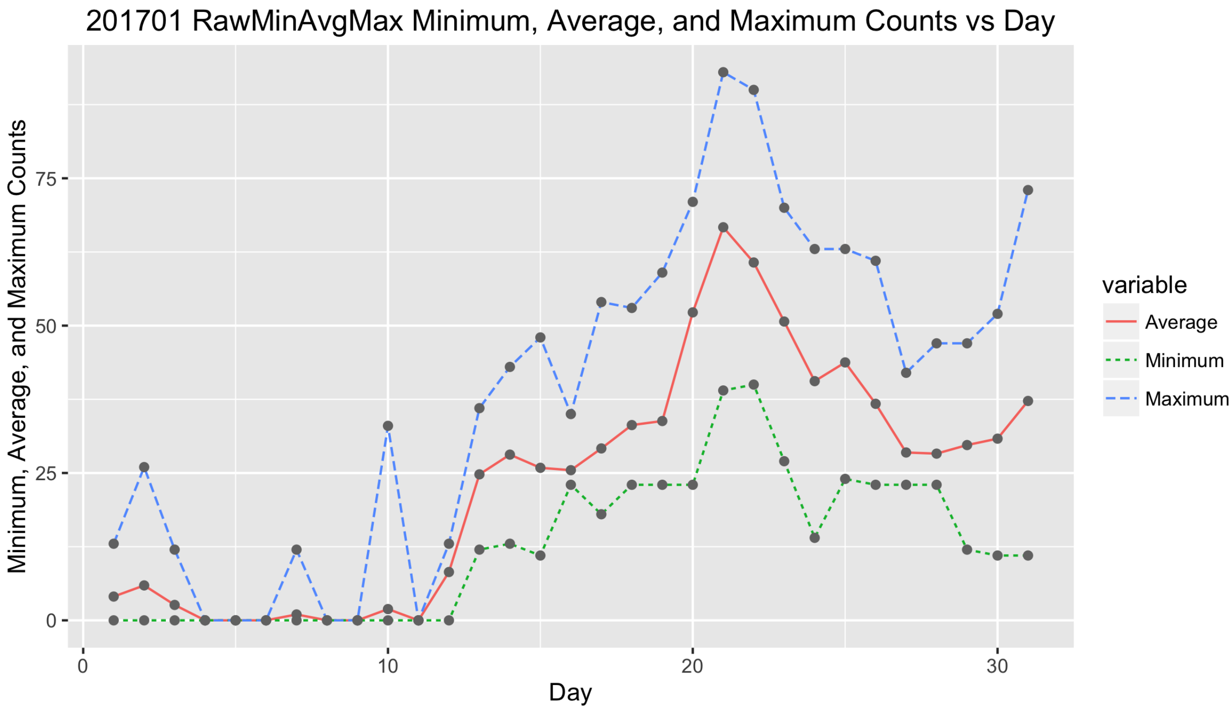

Figure 1: Raw average sunspot count by day of the month.

Table 1: 201701 Daily Raw Counts

Day	Submissions	Minimum	Average	Maximum
1.0000	34.0000	0.0000	4.0370	13.0000
2.0000	29.0000	0.0000	5.9259	26.0000
3.0000	28.0000	0.0000	2.6154	12.0000
4.0000	29.0000	0.0000	0.0000	0.0000
5.0000	29.0000	0.0000	0.0000	0.0000
6.0000	31.0000	0.0000	0.0000	0.0000
7.0000	30.0000	0.0000	1.0000	12.0000
8.0000	29.0000	0.0000	0.0000	0.0000
9.0000	28.0000	0.0000	0.0000	0.0000
10.0000	24.0000	0.0000	1.9130	33.0000
11.0000	24.0000	0.0000	0.0000	0.0000
12.0000	26.0000	0.0000	8.2000	13.0000
13.0000	28.0000	12.0000	24.7692	36.0000
14.0000	31.0000	13.0000	28.1200	43.0000
15.0000	28.0000	11.0000	25.8800	48.0000
16.0000	24.0000	23.0000	25.4762	35.0000
17.0000	26.0000	18.0000	29.1667	54.0000
18.0000	25.0000	23.0000	33.1304	53.0000
19.0000	24.0000	23.0000	33.8182	59.0000
20.0000	26.0000	23.0000	52.2500	71.0000
21.0000	33.0000	39.0000	66.6923	93.0000
22.0000	26.0000	40.0000	60.7143	90.0000
23.0000	25.0000	27.0000	50.6957	70.0000
24.0000	26.0000	14.0000	40.5833	63.0000
25.0000	29.0000	24.0000	43.7600	63.0000
26.0000	27.0000	23.0000	36.7391	61.0000
27.0000	27.0000	23.0000	28.4783	42.0000
28.0000	35.0000	23.0000	28.2857	47.0000
29.0000	35.0000	12.0000	29.7407	47.0000
30.0000	26.0000	11.0000	30.8261	52.0000
31.0000	28.0000	11.0000	37.2222	73.0000

3 Error Tables

Data are for the month of January 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	23.8960	23.3482	24.4439	8.4000	8.7000
2010.06	18.3082	17.8190	18.7973	11.0000	13.6000
2010.07	20.5938	20.1377	21.0499	15.2000	16.1000
2010.08	20.3030	19.8081	20.7979	18.3000	19.6000
2010.09	23.8859	23.3719	24.3998	22.8000	25.2000
2010.10	22.7003	22.2098	23.1909	21.0000	23.5000
2010.11	23.4215	22.8918	23.9513	20.9000	21.6000
2010.12	22.4199	21.7725	23.0673	13.9000	14.5000
2011.01	75.8117	74.1052	77.5183	17.7000	18.7000
2011.02	66.2918	64.7964	67.7872	29.1000	29.6000
2011.03	71.9905	70.5038	73.4773	48.0000	55.8000
2011.04	77.5597	75.8998	79.2196	47.3000	54.4000
2011.05	80.6828	79.0715	82.2942	37.3000	41.5000
2011.06	65.1335	63.7605	66.5066	35.2000	37.0000
2011.07	71.3924	69.8192	72.9655	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	73.7342	72.2879	75.1804	42.4000	50.5000
2011.09	83.9338	82.8284	85.0392	73.8000	78.0000
2011.10	79.1650	77.8039	80.5261	78.9000	88.0000
2011.11	80.4049	78.6789	82.1308	84.6000	96.7000
2011.12	74.7542	73.1086	76.3998	65.8000	73.0000
2012.01	77.8505	76.3037	79.3973	55.8000	58.2000
2012.02	65.8759	64.4731	67.2786	29.2000	33.1000
2012.03	74.4168	73.0849	75.7488	53.1000	64.1000
2012.04	77.2218	74.9392	79.5043	51.4000	55.2000
2012.05	84.5917	83.1132	86.0703	61.8000	69.0000
2012.06	68.1075	66.8982	69.3167	59.7000	64.5000
2012.07	75.6139	74.3382	76.8897	64.2000	51.3000
2012.08	74.3914	73.1449	75.6379	57.7000	63.1000
2012.09	84.6973	83.2469	86.1477	57.7000	61.5000
2012.10	81.4623	79.9230	83.0016	48.3000	53.3000
2012.11	83.7840	82.1041	85.4638	56.7000	61.4000
2012.12	75.6528	74.0604	77.2453	37.4000	40.8000
2013.01	87.8077	86.1654	89.4501	63.8000	62.9000
2013.02	76.0371	74.5546	77.5196	37.8000	38.0000
2013.03	81.7625	80.2219	83.3032	50.6000	57.9000
2013.04	90.0086	88.4931	91.5241	70.6000	72.4000
2013.05	92.4274	90.8333	94.0215	77.4000	78.7000
2013.06	75.0385	73.7019	76.3751	51.0000	52.5000
2013.07	81.2456	79.9799	82.5113	57.0000	57.0000
2013.08	82.0691	80.7866	83.3517	60.0000	66.0000
2013.09	92.5672	90.9676	94.1669	34.6000	36.9000
2013.10	87.4861	85.9331	89.0390	74.5000	85.6000
2013.11	90.0834	88.2113	91.9554	73.9000	77.6000
2013.12	83.4273	81.7521	85.1024	77.8000	90.3000
2014.01	104.2557	102.0648	106.4467	77.4000	82.0000
2014.02	90.2638	88.5501	91.9774	93.9000	102.8000
2014.03	100.5399	98.8428	102.2370	80.9000	92.2000
2014.04	109.3559	107.4972	111.2145	76.9000	84.7000
2014.05	111.8629	110.0716	113.6542	72.3000	75.2000
2014.06	90.8105	89.3383	92.2828	67.2000	71.0000
2014.07	99.7062	98.0806	101.3318	72.5000	72.5000
2014.08	100.1987	98.6918	101.7055	71.2000	74.7000
2014.09	114.1946	112.3526	116.0365	83.2000	87.6000
2014.10	107.6870	105.8842	109.4899	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	111.4246	109.3237	113.5256	65.8000	71.1000
2014.12	100.8632	98.6779	103.0484	75.8000	78.0000
2015.01	63.6071	62.3800	64.8341	65.9000	67.0000
2015.02	55.0705	53.7846	56.3565	42.4000	44.8000
2015.03	60.1280	59.0235	61.2325	38.0000	38.4000
2015.04	66.2088	65.0331	67.3845	49.0000	54.4000
2015.05	67.3268	66.2452	68.4083	56.3000	58.8000
2015.06	55.1235	54.1833	56.0636	50.2000	68.3000
2015.07	59.3182	58.3035	60.3330	47.9000	65.8000
2015.08	61.0782	60.0758	62.0806	39.5000	57.2000
2015.09	69.3103	68.1804	70.4401	49.2000	72.1000
2015.10	65.3315	64.2157	66.4472	39.3000	48.3000
2015.11	68.1086	67.1864	69.0308	39.6000	55.9000
2015.12	61.1567	59.9202	62.3932	36.4000	44.8000
2016.01	35.4977	34.8680	36.1274	33.7000	43.3000
2016.02	30.0788	29.4820	30.6757	38.3000	46.8000
2016.03	32.7355	32.1346	33.3363	30.5000	38.9000
2016.04	35.4247	34.7951	36.0542	26.6000	30.9000
2016.05	37.0174	36.3837	37.6512	33.7000	48.4000
2016.06	29.9344	29.4589	30.4098	13.1000	19.5000
2016.07	32.9029	32.3938	33.4119	21.2000	27.5000
2016.08	33.5277	32.9745	34.0809	33.0000	47.9000
2016.09	37.8140	37.1858	38.4422	27.7000	37.1000
2016.10	35.8884	35.2672	36.5096	22.7000	31.7000
2016.11	37.0025	36.3656	37.6395	14.0000	22.2000
2016.12	33.5405	32.8676	34.2134	11.1000	20.0000
2017.01	23.1652	22.7239	23.6064	18.4000	26.2000

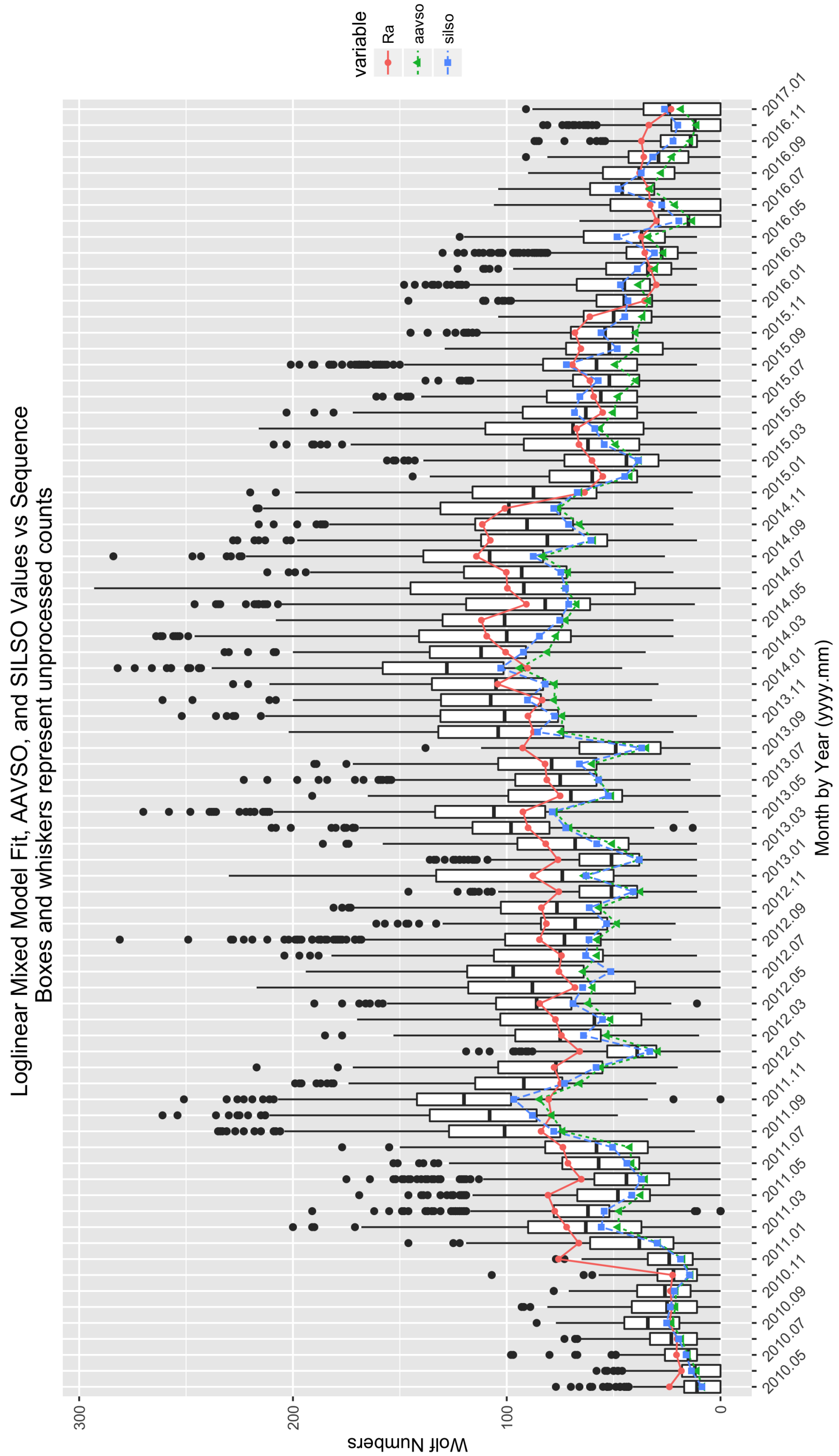

Figure 2: GLMM fitted data for R_a. AAVSO data: https://www.aavso.org/category/tags/solar-bulletin. SILSO data: WDC-SILSO, Royal Observatory of Belgium, Brussels

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: 201701 Parameter Estimates

	Estimate	Std. Error	t-value	Pr(>|t|)
(Intercept)	3.2154	0.0445	72.2104	0.0000
seeF	-0.1894	0.0072	-26.1916	0.0000
seeG	-0.1029	0.0063	-16.3762	0.0000
seeP	-0.2950	0.0106	-27.8181	0.0000
r1000B	-0.0583	0.0831	-0.7009	0.4834
r1500C	0.0365	0.1271	0.2871	0.7741
r2000D	0.0766	0.1549	0.4944	0.6211
r2500E	-0.0011	0.1052	-0.0105	0.9916
r3000F	0.0731	0.1025	0.7135	0.4755
r3500G	0.1228	0.1534	0.8004	0.4235
r5000H	-0.1062	0.2121	-0.5008	0.6165
silsoy	0.1216	0.0738	1.6471	0.0995
year2011	1.2061	0.0154	78.2366	0.0000
year2012	1.2227	0.0154	79.5765	0.0000
year2013	1.3189	0.0153	86.0846	0.0000
year2014	1.5073	0.0152	99.0430	0.0000
year2015	1.0093	0.0156	64.6424	0.0000
year2016	0.4044	0.0166	24.3511	0.0000
year2017	-0.0208	0.0401	-0.5185	0.6041
mon2	-0.1537	0.0121	-12.6802	0.0000
mon3	-0.0649	0.0111	-5.8201	0.0000
mon4	0.0223	0.0112	1.9898	0.0466
mon5	0.0484	0.0106	4.5691	0.0000
mon6	-0.1694	0.0112	-15.1044	0.0000
mon7	-0.0812	0.0108	-7.5408	0.0000
mon8	-0.0653	0.0106	-6.1579	0.0000
mon9	0.0654	0.0102	6.3960	0.0000
mon10	0.0128	0.0108	1.1827	0.2369
mon11	0.0509	0.0110	4.6107	0.0000
mon12	-0.0384	0.0117	-3.2731	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: 201701 Summary of Sunspot Numbers

obs	jd	year	mon	day
ARAG : 2405	Min. :1721096	Min. :2010	Min. : 1.000	Min. : 1.00
CHAG : 2210	1st Qu.:2456004	1st Qu.:2012	1st Qu.: 4.000	1st Qu.: 8.00
BRAB : 2190	Median :2456556	Median :2013	Median : 7.000	Median :16.00
BROB : 1947	Mean :2456262	Mean :2013	Mean : 6.718	Mean :15.72
HOWR : 1817	3rd Qu.:2457170	3rd Qu.:2015	3rd Qu.: 9.000	3rd Qu.:23.00
KNJS : 1809	Max. :2457785	Max. :2017	Max. :12.000	Max. :31.00
(Other):43683				

Table 5: Summary of Sunspot Numbers

see	g	s	w	r	silso
E:10234	Min. : 0.000	Min. : 0.00	Min. : 0.0	0000A :23796	n:37665
F:17222	1st Qu.: 2.000	1st Qu.: 9.00	1st Qu.: 36.0	3000F : 9270	y:18396
G:23962	Median : 4.000	Median : 21.00	Median : 63.0	2500E : 7384	
P: 4643	Mean : 4.295	Mean : 26.74	Mean : 69.7	3500G : 4400	
	3rd Qu.: 6.000	3rd Qu.: 38.00	3rd Qu.: 98.0	1000B : 4074	
	Max. :18.000	Max. :204.00	Max. :293.0	1500C : 3021	
				(Other): 4116	

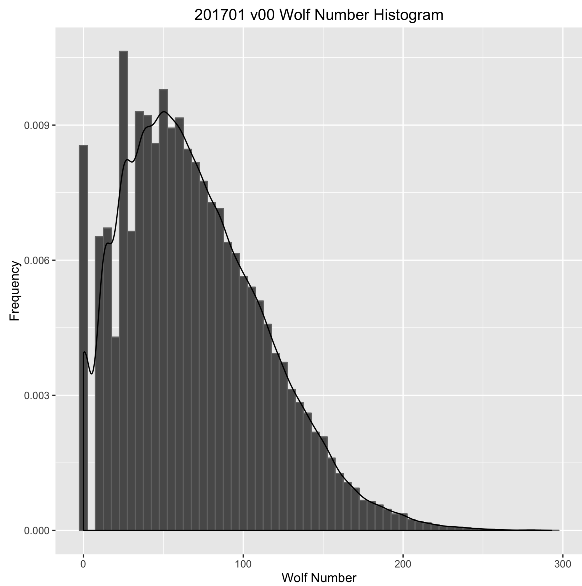

Figure 3: Box plots of raw Wolf number (w) by observer rank.

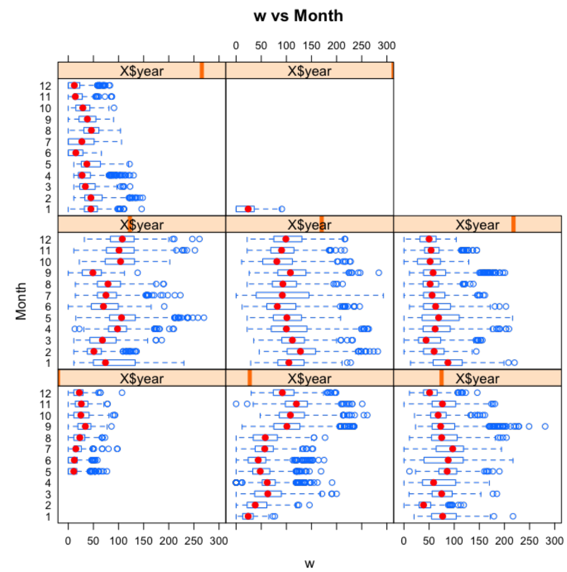

Figure 4: Box plots of raw Wolf number (w) by month and year.

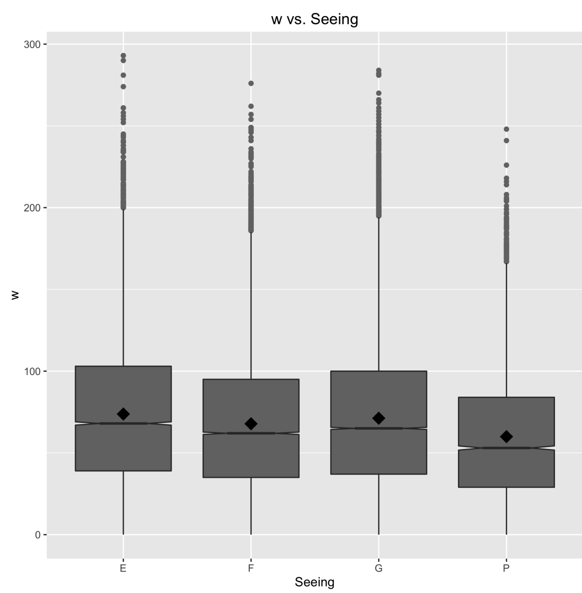

Figure 5: Box plots of raw Wolf number (w) by seeing condition.

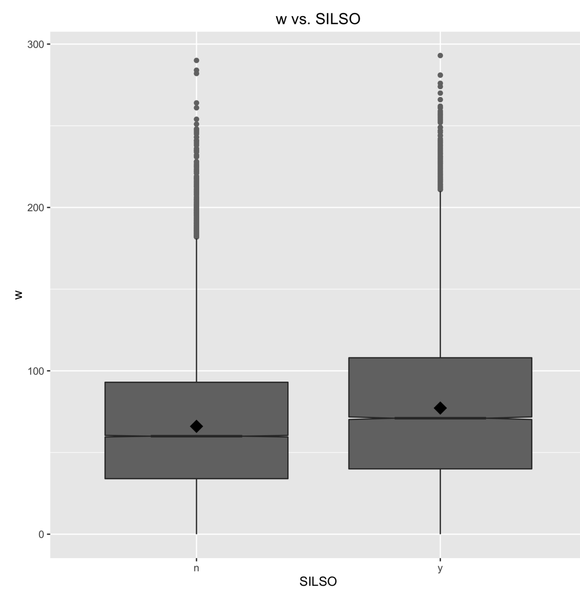

Figure 6: Box plots of raw Wolf number (w) by organization.

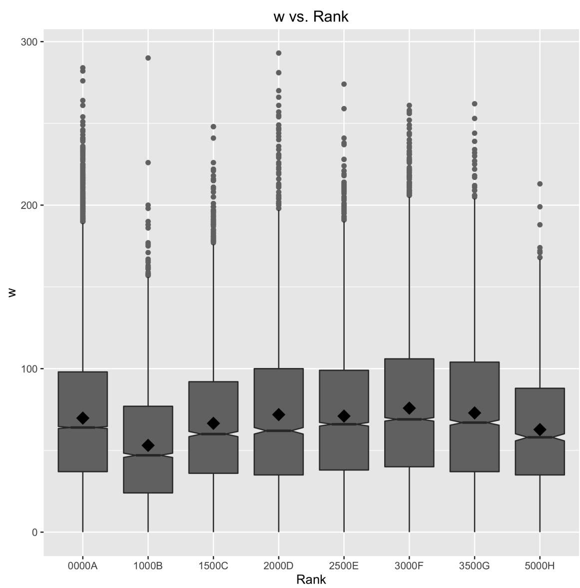

Figure 7: Box plots of raw Wolf number (w) by observer rank.

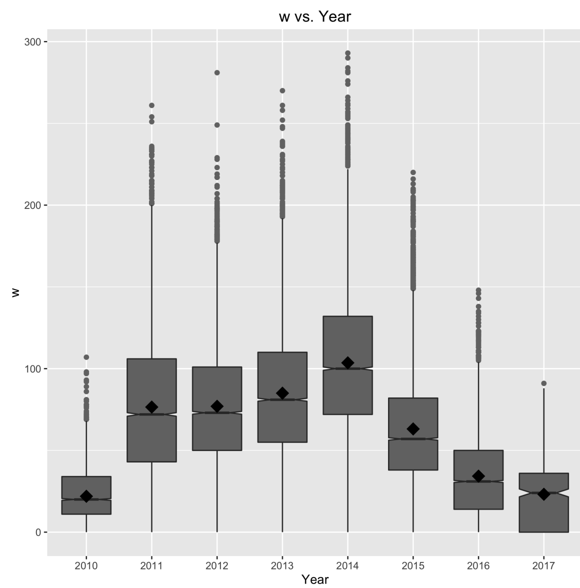

Figure 8: Box plots of raw Wolf number (w) by year.