

Monthly Report (00)

2017.02 Data Set

Tuesday 14th March, 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.

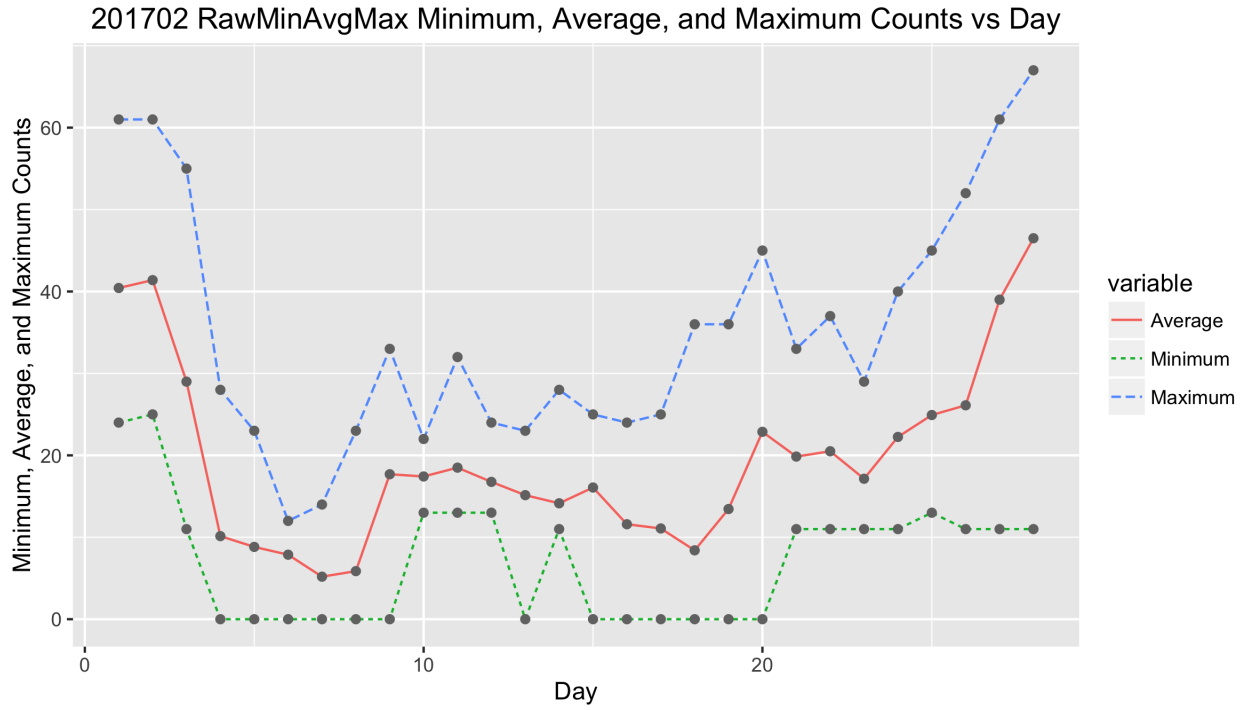


Figure 1: Raw average sunspot count by day of the month.

Table 1: 201702 Daily Raw Counts

Day	Submissions	Minimum	Average	Maximum
1.0000	24.0000	24.0000	40.4286	61.0000
2.0000	22.0000	25.0000	41.3889	61.0000
3.0000	24.0000	11.0000	29.0000	55.0000
4.0000	23.0000	0.0000	10.1429	28.0000
5.0000	24.0000	0.0000	8.8261	23.0000
6.0000	26.0000	0.0000	7.8750	12.0000
7.0000	23.0000	0.0000	5.1905	14.0000
8.0000	27.0000	0.0000	5.8696	23.0000
9.0000	24.0000	0.0000	17.6957	33.0000
10.0000	29.0000	13.0000	17.4231	22.0000
11.0000	19.0000	13.0000	18.5000	32.0000
12.0000	25.0000	13.0000	16.7619	24.0000
13.0000	26.0000	0.0000	15.1429	23.0000
14.0000	34.0000	11.0000	14.1481	28.0000
15.0000	32.0000	0.0000	16.0741	25.0000
16.0000	36.0000	0.0000	11.5926	24.0000
17.0000	28.0000	0.0000	11.0769	25.0000
18.0000	32.0000	0.0000	8.4074	36.0000
19.0000	35.0000	0.0000	13.4483	36.0000
20.0000	27.0000	0.0000	22.8696	45.0000
21.0000	27.0000	11.0000	19.8519	33.0000
22.0000	26.0000	11.0000	20.5000	37.0000
23.0000	23.0000	11.0000	17.1429	29.0000
24.0000	21.0000	11.0000	22.2500	40.0000
25.0000	30.0000	13.0000	24.9231	45.0000
26.0000	35.0000	11.0000	26.1111	52.0000
27.0000	34.0000	11.0000	39.0000	61.0000
28.0000	28.0000	11.0000	46.5000	67.0000

3 Error Tables

Data are for the month of February 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 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	aavso	silso
2010.05	23.8947	23.3466	24.4428	8.4000	8.7000
2010.06	18.3083	17.8189	18.7978	11.0000	13.6000
2010.07	20.5928	20.1365	21.0491	15.2000	16.1000
2010.08	20.3012	19.8062	20.7963	18.3000	19.6000
2010.09	23.8864	23.3721	24.4006	22.8000	25.2000
2010.10	22.7031	22.2123	23.1939	21.0000	23.5000
2010.11	23.4228	22.8929	23.9526	20.9000	21.6000
2010.12	22.4190	21.7711	23.0668	13.9000	14.5000
2011.01	75.8716	74.1627	77.5804	17.7000	18.7000
2011.02	66.2375	64.7426	67.7324	29.1000	29.6000
2011.03	71.9911	70.5035	73.4788	48.0000	55.8000
2011.04	77.5612	75.9007	79.2216	47.3000	54.4000
2011.05	80.6829	79.0708	82.2950	37.3000	41.5000
2011.06	65.1356	63.7619	66.5092	35.2000	37.0000
2011.07	71.3950	69.8211	72.9689	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	73.7383	72.2910	75.1856	42.4000	50.5000
2011.09	83.9304	82.8242	85.0366	73.8000	78.0000
2011.10	79.1648	77.8022	80.5274	78.9000	88.0000
2011.11	80.3991	78.6715	82.1266	84.6000	96.7000
2011.12	74.7482	73.1010	76.3955	65.8000	73.0000
2012.01	77.9086	76.3597	79.4574	55.8000	58.2000
2012.02	65.8133	64.4103	67.2162	29.2000	33.1000
2012.03	74.4119	73.0788	75.7450	53.1000	64.1000
2012.04	77.2201	74.9382	79.5020	51.4000	55.2000
2012.05	84.5949	83.1153	86.0745	61.8000	69.0000
2012.06	68.1076	66.8978	69.3174	59.7000	64.5000
2012.07	75.6165	74.3400	76.8930	64.2000	51.3000
2012.08	74.3912	73.1436	75.6387	57.7000	63.1000
2012.09	84.6998	83.2485	86.1512	57.7000	61.5000
2012.10	81.4627	79.9222	83.0032	48.3000	53.3000
2012.11	83.7828	82.1013	85.4644	56.7000	61.4000
2012.12	75.6561	74.0622	77.2500	37.4000	40.8000
2013.01	87.8701	86.2260	89.5142	63.8000	62.9000
2013.02	75.9746	74.4921	77.4571	37.8000	38.0000
2013.03	81.7665	80.2244	83.3085	50.6000	57.9000
2013.04	90.0078	88.4911	91.5245	70.6000	72.4000
2013.05	92.4256	90.8301	94.0212	77.4000	78.7000
2013.06	75.0401	73.7027	76.3776	51.0000	52.5000
2013.07	81.2452	79.9788	82.5116	57.0000	57.0000
2013.08	82.0703	80.7869	83.3537	60.0000	66.0000
2013.09	92.5659	90.9648	94.1670	34.6000	36.9000
2013.10	87.4828	85.9288	89.0368	74.5000	85.6000
2013.11	90.0827	88.2087	91.9566	73.9000	77.6000
2013.12	83.4262	81.7488	85.1035	77.8000	90.3000
2014.01	104.3290	102.1339	106.5241	77.4000	82.0000
2014.02	90.1811	88.4669	91.8953	93.9000	102.8000
2014.03	100.5440	98.8456	102.2424	80.9000	92.2000
2014.04	109.3587	107.4985	111.2189	76.9000	84.7000
2014.05	111.8630	110.0702	113.6557	72.3000	75.2000
2014.06	90.8110	89.3376	92.2843	67.2000	71.0000
2014.07	99.7043	98.0775	101.3311	72.5000	72.5000
2014.08	100.1987	98.6907	101.7067	71.2000	74.7000
2014.09	114.1972	112.3536	116.0407	83.2000	87.6000
2014.10	107.6861	105.8818	109.4905	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	111.4275	109.3245	113.5305	65.8000	71.1000
2014.12	100.8640	98.6753	103.0528	75.8000	78.0000
2015.01	63.6515	62.4221	64.8808	65.9000	67.0000
2015.02	55.0205	53.7341	56.3069	42.4000	44.8000
2015.03	60.1261	59.0206	61.2315	38.0000	38.4000
2015.04	66.2073	65.0305	67.3840	49.0000	54.4000
2015.05	67.3249	66.2425	68.4074	56.3000	58.8000
2015.06	55.1199	54.1789	56.0609	50.2000	68.3000
2015.07	59.3160	58.3006	60.3314	47.9000	65.8000
2015.08	61.0751	60.0717	62.0785	39.5000	57.2000
2015.09	69.3126	68.1823	70.4429	49.2000	72.1000
2015.10	65.3342	64.2180	66.4503	39.3000	48.3000
2015.11	68.1110	67.1882	69.0339	39.6000	55.9000
2015.12	61.1601	59.9230	62.3972	36.4000	44.8000
2016.01	35.5237	34.8928	36.1546	33.7000	43.3000
2016.02	30.0539	29.4567	30.6510	38.3000	46.8000
2016.03	32.7349	32.1332	33.3367	30.5000	38.9000
2016.04	35.4232	34.7930	36.0535	26.6000	30.9000
2016.05	37.0182	36.3839	37.6526	33.7000	48.4000
2016.06	29.9339	29.4580	30.4098	13.1000	19.5000
2016.07	32.9033	32.3938	33.4127	21.2000	27.5000
2016.08	33.5275	32.9737	34.0812	33.0000	47.9000
2016.09	37.8151	37.1863	38.4439	27.7000	37.1000
2016.10	35.8877	35.2657	36.5097	22.7000	31.7000
2016.11	37.0021	36.3642	37.6400	14.0000	22.2000
2016.12	33.5412	32.8672	34.2151	11.1000	20.0000
2017.01	22.8269	22.3915	23.2623	18.4000	26.2000
2017.02	19.1431	18.7680	19.5182	14.4000	20.6000

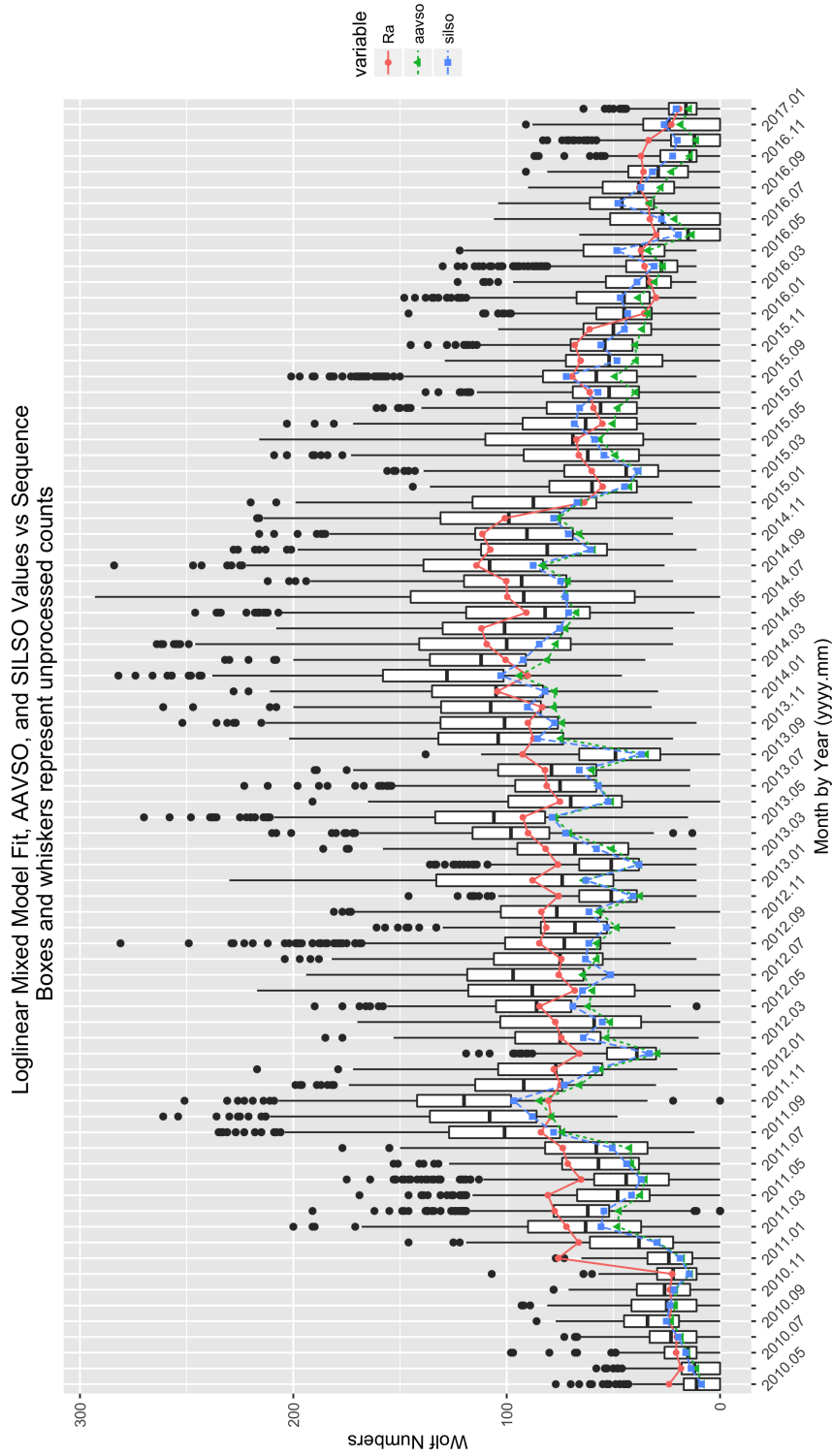


Figure 2: GLMM fitted data for R_a . AAVSO data: <https://www.aavso.org/category/tags/solar-bulletin>. SILSO data: WDC-SILSO, Royal Observatory of Belgium, Brussels

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: 201702 Parameter Estimates

	Estimate	Std. Error	t-value	$Pr(> t)$
(Intercept)	3.2159	0.0445	72.2465	0.0000
seeF	-0.1894	0.0072	-26.2799	0.0000
seeG	-0.1030	0.0063	-16.4412	0.0000
seeP	-0.2945	0.0106	-27.8632	0.0000
r1000B	-0.0581	0.0832	-0.6989	0.4846
r1500C	0.0366	0.1271	0.2875	0.7737
r2000D	0.0766	0.1550	0.4942	0.6212
r2500E	-0.0008	0.1052	-0.0077	0.9938
r3000F	0.0735	0.1025	0.7175	0.4731
r3500G	0.1229	0.1534	0.8008	0.4233
r5000H	-0.1057	0.2122	-0.4982	0.6183
silsoy	0.1217	0.0738	1.6482	0.0993
year2011	1.2060	0.0154	78.4010	0.0000
year2012	1.2227	0.0153	79.7467	0.0000
year2013	1.3189	0.0153	86.2691	0.0000
year2014	1.5074	0.0152	99.2557	0.0000
year2015	1.0093	0.0156	64.7807	0.0000
year2016	0.4044	0.0166	24.4050	0.0000
year2017	-0.0362	0.0318	-1.1404	0.2541
mon2	-0.1552	0.0118	-13.1366	0.0000
mon3	-0.0655	0.0111	-5.9289	0.0000
mon4	0.0216	0.0111	1.9436	0.0519
mon5	0.0477	0.0105	4.5400	0.0000
mon6	-0.1701	0.0111	-15.2862	0.0000
mon7	-0.0819	0.0107	-7.6687	0.0000
mon8	-0.0660	0.0105	-6.2759	0.0000
mon9	0.0647	0.0101	6.3879	0.0000
mon10	0.0121	0.0107	1.1271	0.2597
mon11	0.0502	0.0110	4.5836	0.0000
mon12	-0.0391	0.0116	-3.3559	0.0008

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: 201702 Summary of Sunspot Numbers

obs	jd	year	mon	day
ARAG : 2432	Min. :1721096	Min. :2010	Min. : 1.000	Min. : 1.00
CHAG : 2233	1st Qu.:2456009	1st Qu.:2012	1st Qu.: 4.000	1st Qu.: 8.00
BRAB : 2217	Median :2456566	Median :2013	Median : 7.000	Median :16.00
BROB : 1962	Mean :2456276	Mean :2013	Mean : 6.675	Mean :15.72
HOWR : 1838	3rd Qu.:2457186	3rd Qu.:2015	3rd Qu.: 9.000	3rd Qu.:23.00
KNJS : 1836	Max. :2457813	Max. :2017	Max. :12.000	Max. :31.00
(Other):44056				

Table 5: Summary of Sunspot Numbers

see	g	s	w	r	silso
E:10287	Min. : 0.000	Min. : 0.00	Min. : 0.00	0000A :24007	n:38001
F:17422	1st Qu.: 2.000	1st Qu.: 9.00	1st Qu.: 36.00	3000F : 9376	y:18573
G:24163	Median : 4.000	Median : 21.00	Median : 63.00	2500E : 7468	
P: 4702	Mean : 4.269	Mean : 26.55	Mean : 69.24	3500G : 4450	
	3rd Qu.: 6.000	3rd Qu.: 38.00	3rd Qu.: 97.00	1000B : 4092	
	Max. :18.000	Max. :204.00	Max. :293.00	1500C : 3021	
				(Other): 4160	

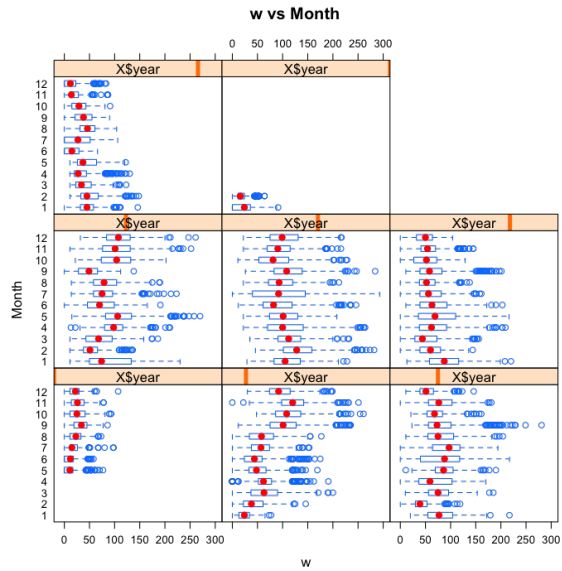
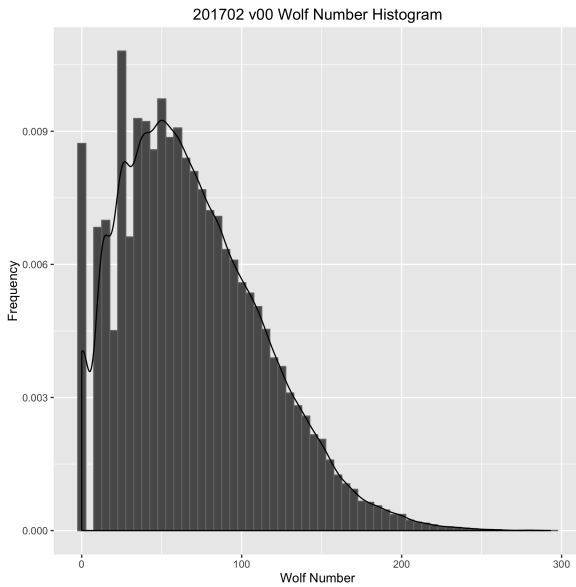


Figure 3: Box plots of raw Wolf number (w) by observer rank.

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

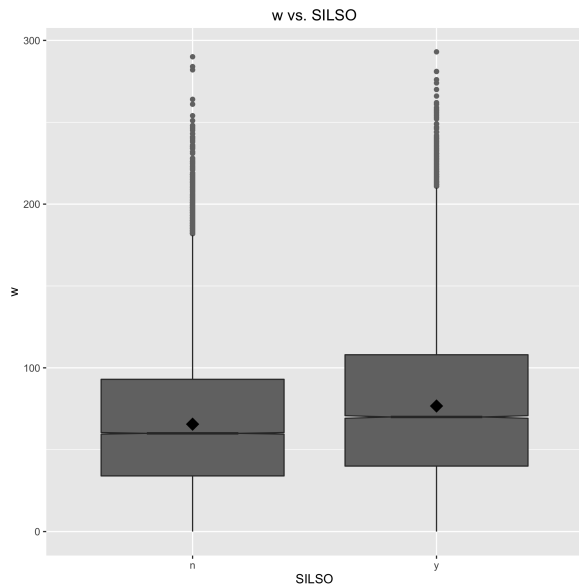
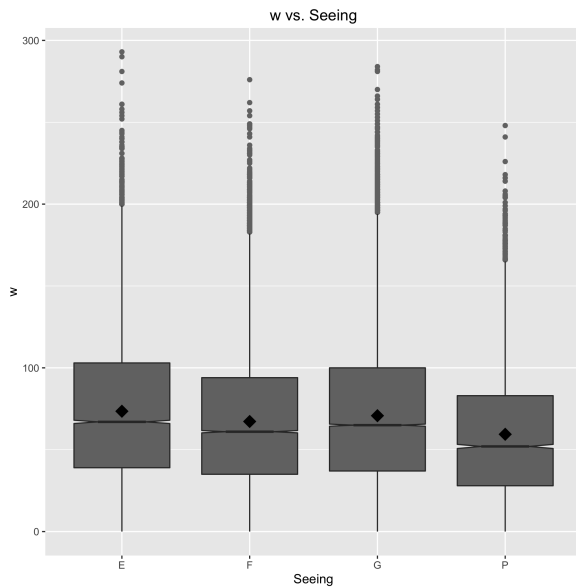


Figure 5: Box plots of raw Wolf number (w) by seeing condition.

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

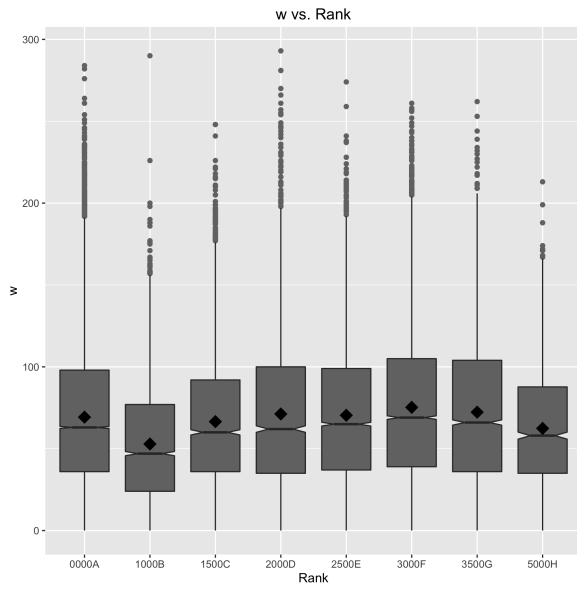


Figure 7: Box plots of raw Wolf number (w) by observer rank.

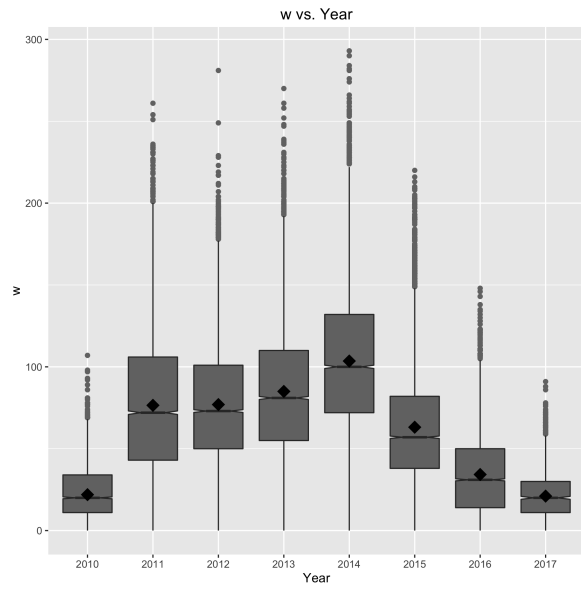


Figure 8: Box plots of raw Wolf number (w) by year.