

## Data 1: Bernoulli, 20news

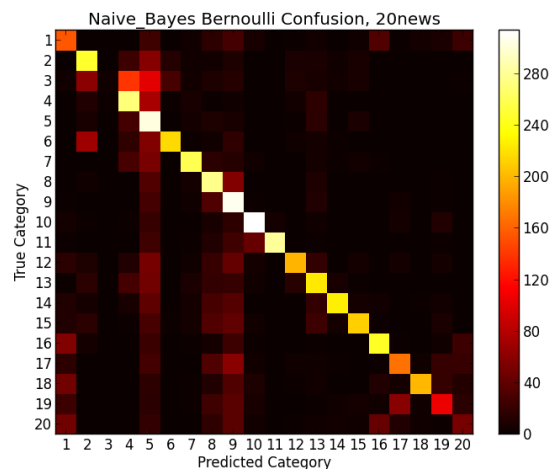
Report File: Naive\_Bayes\_Bernoulli\_20news\_report.txt

### Classification Report:

	precision	recall	f1-score	support
alt.atheism	0.39	0.49	0.43	319
comp.graphics	0.54	0.64	0.59	389
comp.os.ms-windows.misc	0.38	0.01	0.01	394
comp.sys.ibm.pc.hardware	0.48	0.69	0.57	392
comp.sys.mac.hardware	0.29	0.79	0.42	385
comp.windows.x	0.82	0.55	0.66	395
misc.forsale	0.80	0.67	0.73	390
rec.autos	0.48	0.70	0.57	396
rec.motorcycles	0.37	0.78	0.50	398
rec.sport.baseball	0.77	0.79	0.78	397
rec.sport.hockey	0.97	0.70	0.82	399
sci.crypt	0.79	0.50	0.61	396
sci.electronics	0.58	0.57	0.58	393
sci.med	0.86	0.57	0.68	396
sci.space	0.79	0.54	0.64	394
soc.religion.christian	0.70	0.62	0.66	398
talk.politics.guns	0.63	0.46	0.53	364
talk.politics.mideast	0.91	0.53	0.67	376
talk.politics.misc	0.54	0.34	0.42	310
talk.religion.misc	0.33	0.20	0.25	251
avg / total	0.63	0.57	0.56	7532

### Confusion Matrix:

155	1	0	1	25	0	4	16	27	7	0	2	4	1	3	34	2	5	8	24
1	248	0	24	58	13	3	4	9	0	0	8	8	3	7	1	0	1	1	0
3	61	3	139	103	29	4	9	13	0	0	9	7	4	7	1	0	0	0	2
0	10	2	272	74	2	7	2	3	0	0	4	16	0	0	0	0	0	0	0
0	7	1	27	303	0	6	9	7	0	0	2	15	0	8	0	0	0	0	0
0	69	1	18	56	217	2	3	18	1	0	3	5	2	0	0	0	0	0	0
0	1	0	29	52	1	260	15	13	3	1	0	6	2	4	2	0	1	0	0
1	3	0	1	34	0	6	276	57	1	0	1	9	1	1	1	1	0	2	1
2	1	0	0	28	0	3	33	309	2	0	1	11	0	0	0	4	1	2	1
6	2	0	2	22	0	1	7	17	314	6	0	3	1	0	1	5	0	10	0
2	1	1	0	25	0	2	10	25	44	281	1	2	1	1	0	1	0	2	0
15	9	0	10	51	1	3	21	42	5	0	199	19	1	6	0	6	1	6	1
2	15	0	27	49	1	8	19	21	3	0	14	224	7	2	0	0	0	1	0
11	5	0	7	40	0	5	29	35	1	0	0	19	226	6	4	1	2	4	1
10	15	0	2	29	2	5	34	41	3	0	0	24	5	212	2	1	1	8	0
56	6	0	1	23	0	2	6	24	2	0	0	2	1	0	246	2	0	3	24
17	0	0	1	27	0	1	32	59	3	1	3	3	2	1	2	167	3	21	21
48	0	0	1	16	0	2	10	34	9	0	2	4	0	1	10	6	201	20	12
24	1	0	1	15	0	2	25	37	7	0	1	2	3	6	3	59	3	106	15
49	1	0	1	19	0	0	18	38	4	0	2	3	4	3	43	10	3	2	51



Data 2: Bernoulli, 20news5 and 20news4

Report File: Naive\_Bayes\_Bernoulli\_20news5\_report.txt

-----

Classification Report:

	precision	recall	f1-score	support
computers	0.94	0.76	0.84	1955
recreation	0.55	0.96	0.70	1590
science	0.78	0.62	0.69	1579
politics	0.85	0.52	0.65	1050
religion	0.77	0.70	0.73	968
avg / total	0.78	0.73	0.73	7142

-----

Confusion Matrix:

1489	297	158	1	10
3	1524	35	20	8
90	439	980	34	36
1	291	56	548	154
9	217	23	39	680

Report File: Naive\_Bayes\_Bernoulli\_20news4\_report.txt

-----

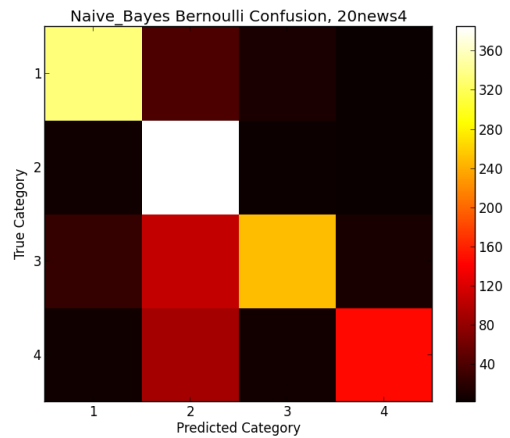
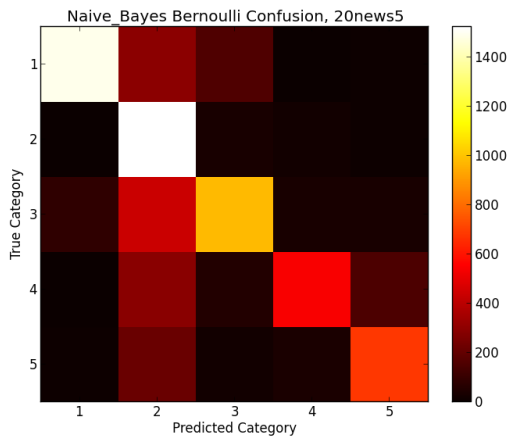
Classification Report:

	precision	recall	f1-score	support
comp.graphics	0.90	0.86	0.88	389
talk.religion.misc	0.62	0.97	0.75	396
sci.space	0.92	0.63	0.75	394
rec.autos	0.91	0.59	0.71	251
avg / total	0.83	0.78	0.78	1430

-----

Confusion Matrix:

334	40	12	3
5	385	4	2
27	108	249	10
6	91	7	147



Data 3: Multinomial, 20news

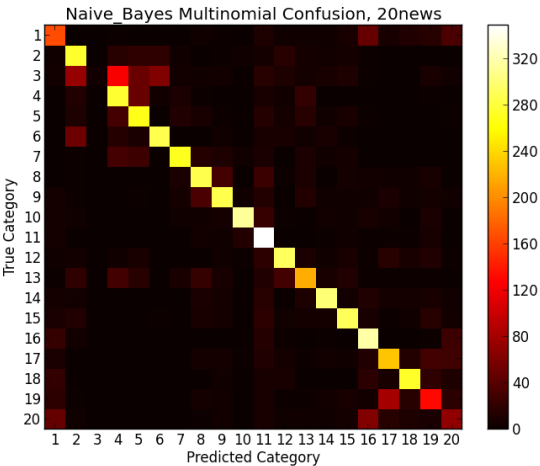
Report File: Naive\_Bayes\_Multinomial\_20news\_report.txt

Classification Report:

	precision	recall	f1-score	support
alt.atheism	0.50	0.52	0.51	319
comp.graphics	0.57	0.71	0.63	389
comp.os.ms-windows.misc	0.71	0.01	0.02	394
comp.sys.ibm.pc.hardware	0.52	0.71	0.60	392
comp.sys.mac.hardware	0.59	0.70	0.64	385
comp.windows.x	0.75	0.73	0.74	395
misc.forsale	0.82	0.69	0.75	390
rec.autos	0.69	0.72	0.71	396
rec.motorcycles	0.72	0.72	0.72	398
rec.sport.baseball	0.90	0.79	0.84	397
rec.sport.hockey	0.59	0.87	0.71	399
sci.crypt	0.73	0.74	0.74	396
sci.electronics	0.64	0.55	0.59	393
sci.med	0.80	0.76	0.78	396
sci.space	0.75	0.74	0.74	394
soc.religion.christian	0.62	0.80	0.70	398
talk.politics.guns	0.59	0.63	0.61	364
talk.politics.mideast	0.78	0.73	0.75	376
talk.politics.misc	0.46	0.43	0.44	310
talk.religion.misc	0.34	0.28	0.31	251
avg / total	0.66	0.65	0.64	7532

Confusion Matrix:

167	0	0	2	0	1	0	3	2	1	10	4	4	3	7	49	7	11	15	33
4	276	0	16	20	20	4	1	2	4	6	16	6	5	6	2	0	0	1	0
5	73	5	125	50	62	5	3	5	0	15	10	6	7	10	2	0	0	7	4
0	12	0	278	50	4	9	2	1	1	7	4	21	1	0	0	0	0	2	0
1	10	1	30	268	1	11	8	2	1	14	6	16	4	7	2	1	0	0	2
0	51	1	14	9	287	1	1	3	0	8	7	4	7	2	0	0	0	0	0
0	2	0	31	27	0	271	14	10	4	9	1	9	3	6	1	0	1	1	0
1	2	0	1	1	0	9	287	31	1	25	2	9	1	6	5	4	3	7	1
6	2	0	0	2	1	6	33	287	3	13	0	13	3	3	4	9	3	6	4
6	3	0	0	0	1	4	4	6	313	24	2	1	5	4	8	5	1	9	1
6	0	0	1	0	0	0	5	4	13	349	1	1	2	4	2	2	2	7	0
3	5	0	3	7	1	1	1	4	2	18	293	9	3	8	2	15	7	12	2
2	20	0	29	16	0	7	23	8	1	11	29	218	8	12	2	2	2	2	1
6	5	0	1	0	0	1	9	6	1	15	1	10	301	5	12	6	6	8	3
9	11	0	1	0	2	0	8	6	0	18	3	6	6	291	7	2	3	16	5
22	5	0	1	1	0	0	0	1	1	14	2	0	2	0	318	1	1	2	27
9	0	0	0	0	0	1	6	6	2	11	6	2	5	4	12	230	13	29	28
21	1	0	1	0	1	1	1	4	1	7	7	1	0	0	17	8	275	20	10
19	2	0	0	0	1	1	6	5	0	8	2	2	5	8	5	81	16	132	17
48	4	0	1	0	0	0	2	5	0	7	3	4	3	6	63	16	10	9	70



Data 4: Multinomial, 20news5 and 20news4

Report File: Naive\_Bayes\_Multinomial\_20news5\_report.txt

Classification Report:

	precision	recall	f1-score	support
computers	0.86	0.75	0.80	1955
recreation	0.92	0.85	0.88	1590
science	0.68	0.82	0.74	1579
politics	0.77	0.76	0.77	1050
religion	0.77	0.82	0.79	968
avg / total	0.81	0.80	0.80	7142

Confusion Matrix:

1461	44	414	13	23
71	1351	80	50	38
100	40	1296	95	48
28	24	67	798	133
33	17	50	76	792

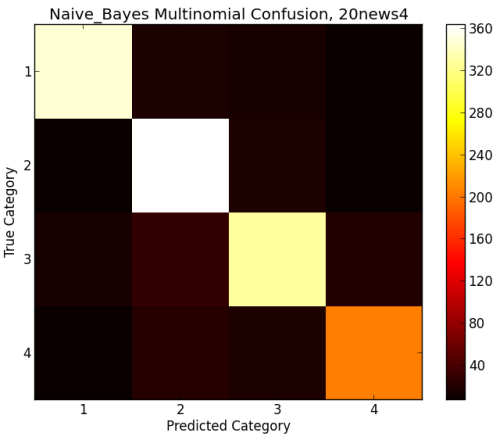
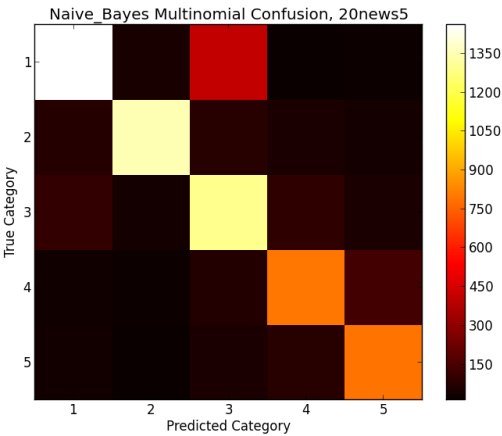
Report File: Naive\_Bayes\_Multinomial\_20news4\_report.txt

Classification Report:

	precision	recall	f1-score	support
comp.graphics	0.92	0.89	0.91	389
talk.religion.misc	0.84	0.92	0.88	396
sci.space	0.87	0.84	0.85	394
rec.autos	0.85	0.82	0.84	251
avg / total	0.87	0.87	0.87	1430

Confusion Matrix:

348	18	15	8
8	364	17	7
15	29	330	20
7	21	18	205



## Data 5: TFIDF, 20news

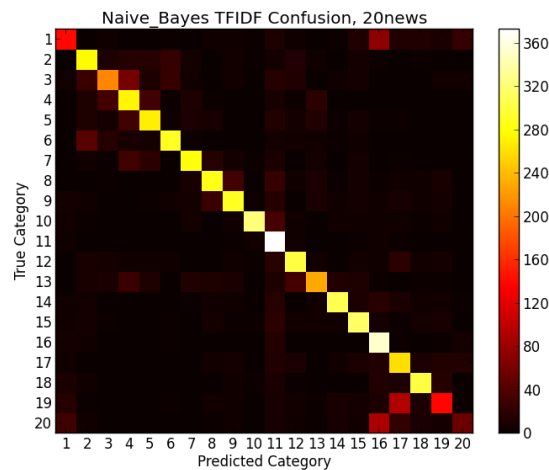
Report File: Naive\_Bayes\_TFIDF\_20news\_report.txt

### Classification Report:

	precision	recall	f1-score	support
alt.atheism	0.59	0.44	0.51	319
comp.graphics	0.66	0.71	0.68	389
comp.os.ms-windows.misc	0.69	0.53	0.60	394
comp.sys.ibm.pc.hardware	0.60	0.70	0.65	392
comp.sys.mac.hardware	0.73	0.70	0.71	385
comp.windows.x	0.80	0.74	0.77	395
misc.forsale	0.80	0.72	0.76	390
rec.autos	0.75	0.72	0.74	396
rec.motorcycles	0.75	0.73	0.74	398
rec.sport.baseball	0.93	0.81	0.87	397
rec.sport.hockey	0.59	0.93	0.72	399
sci.crypt	0.73	0.77	0.75	396
sci.electronics	0.72	0.58	0.65	393
sci.med	0.84	0.78	0.81	396
sci.space	0.74	0.80	0.77	394
soc.religion.christian	0.59	0.89	0.71	398
talk.politics.guns	0.58	0.71	0.64	364
talk.politics.mideast	0.82	0.81	0.81	376
talk.politics.misc	0.60	0.45	0.51	310
talk.religion.misc	0.50	0.22	0.30	251
avg / total	0.71	0.70	0.70	7532

### Confusion Matrix:

141	0	3	1	1	2	0	3	3	2	11	4	1	2	12	74	12	12	10	25
1	276	11	15	15	24	6	0	5	2	7	13	3	0	7	2	0	2	0	0
3	27	210	59	11	25	5	1	5	0	16	12	1	3	7	1	0	1	4	3
0	11	31	273	32	2	11	2	0	0	8	2	18	0	1	0	0	0	1	0
0	11	7	30	269	4	9	8	1	0	14	7	13	2	6	2	2	0	0	0
0	45	16	8	7	291	2	0	0	1	7	5	5	2	4	1	1	0	0	0
1	4	2	30	19	0	281	16	6	2	10	1	7	1	6	1	2	0	1	0
1	1	2	1	1	0	8	286	32	0	25	3	10	2	7	2	4	3	8	0
7	3	1	0	2	3	6	27	291	2	15	0	10	3	6	4	8	3	6	1
5	2	0	0	0	1	6	0	3	323	34	3	1	3	3	3	4	2	4	0
4	0	0	0	0	1	0	1	2	5	373	3	0	1	2	5	2	0	0	0
1	8	7	3	3	3	1	0	4	2	17	303	3	1	5	3	19	5	7	1
1	8	9	27	11	1	12	11	10	1	11	33	229	12	11	2	0	2	2	0
4	5	0	1	0	0	2	5	3	1	16	0	5	308	10	17	9	3	5	2
4	6	2	1	0	2	1	6	2	2	18	3	5	4	316	6	2	6	8	0
7	3	1	0	0	2	0	1	1	1	14	1	0	1	1	354	2	0	2	7
4	0	0	1	0	1	1	5	7	1	12	10	1	4	10	12	260	9	14	12
10	3	1	1	0	1	0	1	4	2	10	2	1	0	0	13	9	303	15	0
15	2	0	0	0	2	1	5	3	0	8	5	2	8	7	9	92	10	139	2
30	4	1	1	0	0	0	2	4	0	8	4	2	8	5	89	23	9	7	54



## Data 6: TFIDF, 20news5 and 20news4

Report File: Naive\_Bayes\_TFIDF\_20news5\_report.txt

### Classification Report:

	precision	recall	f1-score	support
computers	0.83	0.92	0.87	1955
recreation	0.91	0.87	0.89	1590
science	0.81	0.77	0.79	1579
politics	0.80	0.79	0.79	1050
religion	0.83	0.80	0.81	968
avg / total	0.84	0.84	0.84	7142

### Confusion Matrix:

1796	22	116	11	10
86	1384	67	31	22
193	61	1213	81	31
38	35	56	826	95
52	21	43	82	770

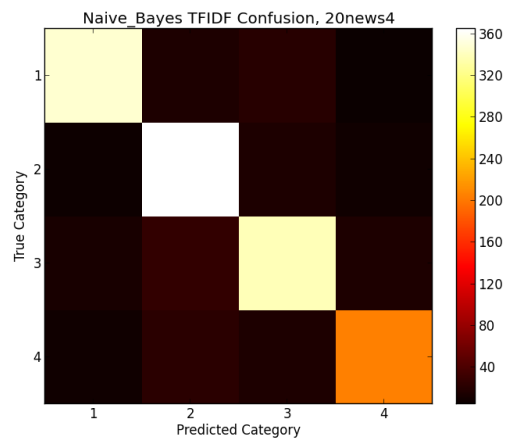
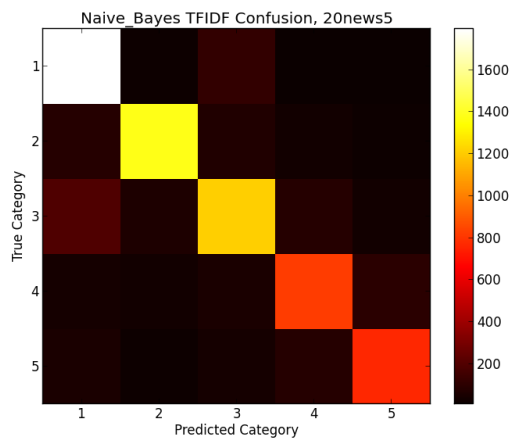
Report File: Naive\_Bayes\_TFIDF\_20news4\_report.txt

### Classification Report:

	precision	recall	f1-score	support
comp.graphics	0.93	0.89	0.91	389
talk.religion.misc	0.85	0.92	0.88	396
sci.space	0.87	0.86	0.86	394
rec.autos	0.88	0.82	0.85	251
avg / total	0.88	0.88	0.88	1430

### Confusion Matrix:

348	15	21	5
7	365	15	9
13	27	339	15
8	22	16	205



Data 7: Bernoulli, Reuters

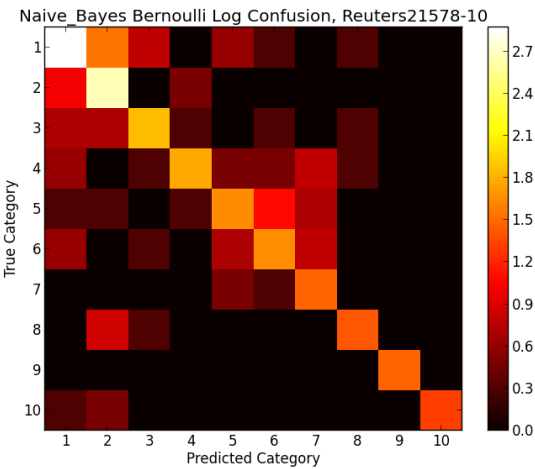
Report File: Naive\_Bayes\_Bernoulli\_Reuters21578-10\_report.txt

Classification Report:

	precision	recall	f1-score	support
earn	0.97	0.94	0.96	789
acq	0.91	0.98	0.94	472
crude	0.90	0.87	0.88	82
trade	0.94	0.81	0.87	72
money-fx	0.80	0.70	0.75	61
interest	0.73	0.77	0.75	57
money-supply	0.67	0.91	0.77	32
ship	0.89	0.78	0.83	32
sugar	1.00	1.00	1.00	29
coffee	1.00	0.87	0.93	23
avg / total	0.93	0.92	0.92	1649

Confusion Matrix:

745	34	5	0	3	1	0	1	0	0
9	461	0	2	0	0	0	0	0	0
4	4	71	1	0	1	0	1	0	0
3	0	1	58	2	2	5	1	0	0
1	1	0	1	43	11	4	0	0	0
3	0	1	0	4	44	5	0	0	0
0	0	0	0	2	1	29	0	0	0
0	6	1	0	0	0	0	25	0	0
0	0	0	0	0	0	0	0	29	0
1	2	0	0	0	0	0	0	0	20



Data 8: Multinomial, Reuters

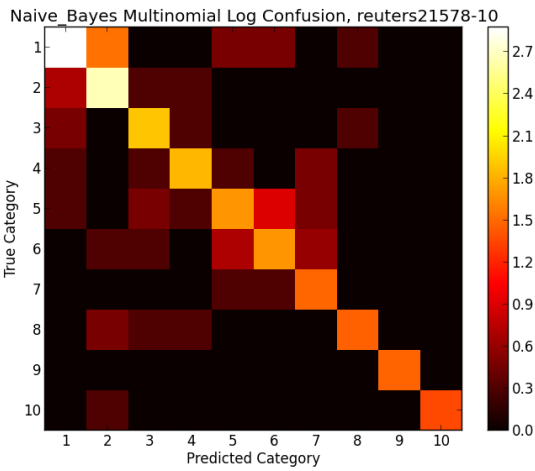
Report File: Naive\_Bayes\_Multinomial\_reuters21578-10\_report.txt

Classification Report:

	precision	recall	f1-score	support
earn	0.99	0.95	0.97	789
acq	0.93	0.99	0.96	472
crude	0.93	0.95	0.94	82
trade	0.94	0.93	0.94	72
money-fx	0.86	0.79	0.82	61
interest	0.83	0.84	0.83	57
money-supply	0.81	0.94	0.87	32
ship	0.93	0.88	0.90	32
sugar	1.00	1.00	1.00	29
coffee	1.00	0.96	0.98	23
avg / total	0.95	0.95	0.95	1649

Confusion Matrix:

751	33	0	0	2	2	0	1	0	0
4	466	1	1	0	0	0	0	0	0
2	0	78	1	0	0	0	1	0	0
1	0	1	67	1	0	2	0	0	0
1	0	2	1	48	7	2	0	0	0
0	1	1	0	4	48	3	0	0	0
0	0	0	0	1	1	30	0	0	0
0	2	1	1	0	0	0	28	0	0
0	0	0	0	0	0	0	0	29	0
0	1	0	0	0	0	0	0	0	22





Data 9: TFIDF, Reuters

Report File: Naive\_Bayes\_TFIDF\_reuters21578-10\_report.txt

Classification Report:

	precision	recall	f1-score	support
earn	0.98	0.95	0.97	789
acq	0.92	0.99	0.95	472
crude	0.93	0.96	0.95	82
trade	0.92	0.99	0.95	72
money-fx	0.88	0.84	0.86	61
interest	0.89	0.84	0.86	57
money-supply	0.94	0.91	0.92	32
ship	0.96	0.81	0.88	32
sugar	1.00	1.00	1.00	29
coffee	1.00	0.91	0.95	23
avg / total	0.95	0.95	0.95	1649

Confusion Matrix:

750	34	0	1	1	2	0	1	0	0
5	465	1	1	0	0	0	0	0	0
2	0	79	1	0	0	0	0	0	0
1	0	0	71	0	0	0	0	0	0
4	0	2	2	51	2	0	0	0	0
0	1	1	0	5	48	2	0	0	0
0	0	0	0	1	2	29	0	0	0
1	2	2	1	0	0	0	26	0	0
0	0	0	0	0	0	0	0	29	0
0	2	0	0	0	0	0	0	0	21

