

HW1

- Basic idea of hypothesis testing
- ANOVA
- Basic data analysis skills, visualization
- Dealing with real data
- Sampling
- Regression



Category	Theme	No.
1	Online Game	484
2	School Alumni	300
3	House & Living	196
4	Stock Market	425
5	Organizations & Industry	635

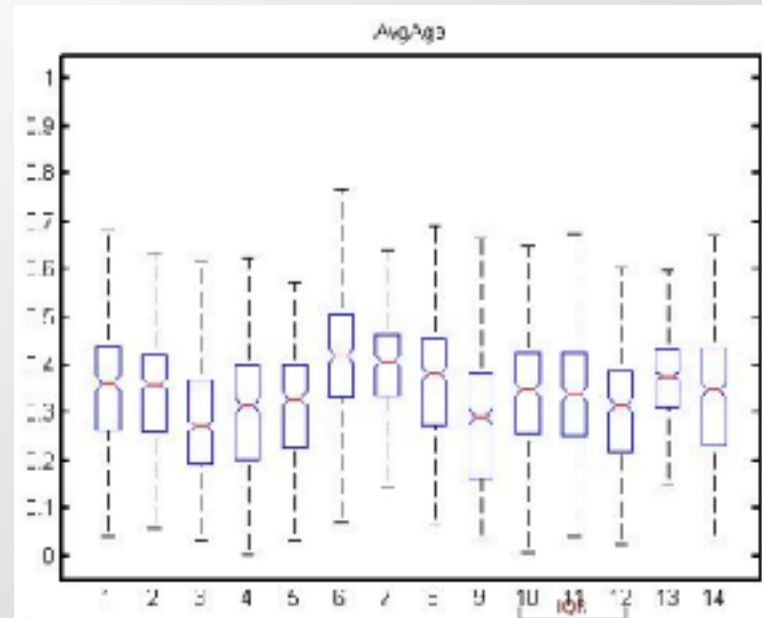
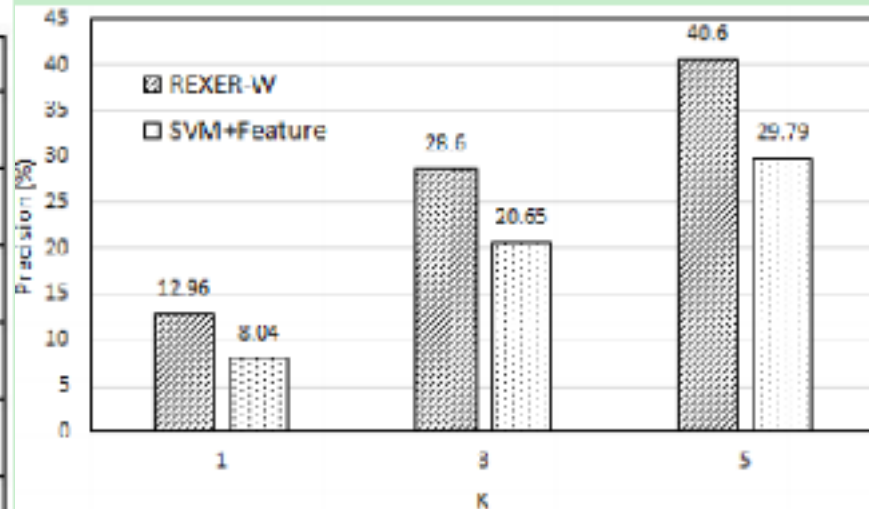
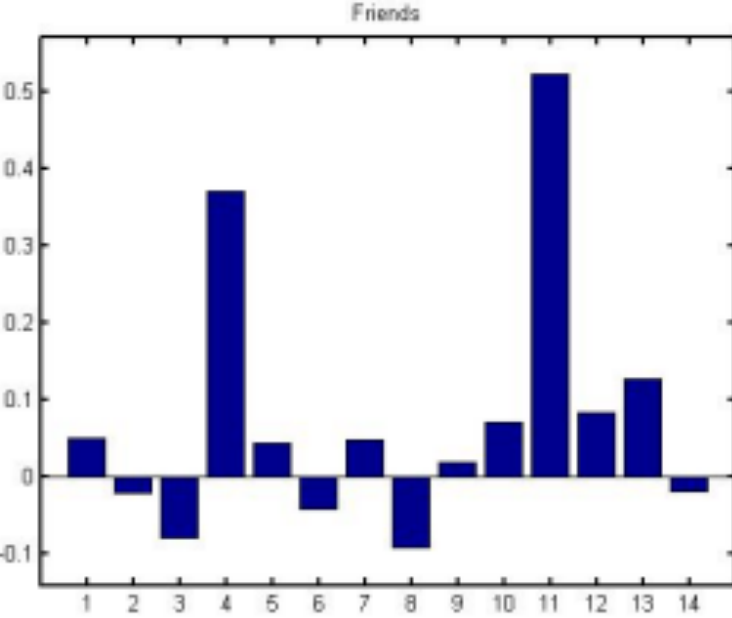
群名	群类别	群人数	消息数	稠密度	性别比	平均年龄	年龄差	地域集中度	手机比例	会话数	无回应比例	夜聊比例	图片比例
黑手酒吧	1	63	315	0.051715	0.882353	26.68182	6.409413	0.222222	0.088889	23	0.434783	0.085714	0.069841
领秀嘉园	1	73	2	0.010274	0.57971	27.5	9.042524	0.40678	0	2	1	1	0
丶~3.無黑\導、1.群	1	73	4	0.024163	0.75	23.41538	4.699893	0.287879	0.5	4	1	0	0.25
上海网管工作联系基地	1	91	14	0.006105	0.901639	29.40909	2.854965	0.340426	0.142857	7	1	0.25	0
tbsgame.net/bbs	1	133	2112	0.011107	0.961832	30.31967	4.669866	0.110236	0.16714	327	0.53211	0.10464	0.037879
迪拜DOTA粉丝群	1	85	91	0.035014	0.903614	26.1125	4.415863	0.617284	0.153846	30	0.7	0.142857	0.065934
血踏联盟	1	60	22	0.038136	0.807692	29.7381	4.619109	0.357143	1	4	0.5	1	0.045455
黑手军团	1	74	115	0.086264	0.833333	26.31343	4.229283	0.228571	0.208696	20	0.55	0.034783	0.026087
百度赤水贴吧	1	55	1940	0.091919	0.759259	24.41509	4.651776	0.745455	0.548969	193	0.414508	0.238783	0.017526

Feature can distinguish categories? => features varied significantly across the categories?

Feature like Age follows Normality. How about other behavior and social features?

- Tools: Matlab, R, python, spss, excel...
- Detailed problem description, see homework1.pdf
 - Problem definition
 - Understand data and Verify assumptions
 - Experiment.
 - Visualization
 - Caution: Real data v.s. Ideal data
 - DDL: November 15th, 2018
- Hand in: An experiment reports and codes
- For more details and beyond:
 - <http://pengcui.thumedia lab.com>





2-D=>contain more information

