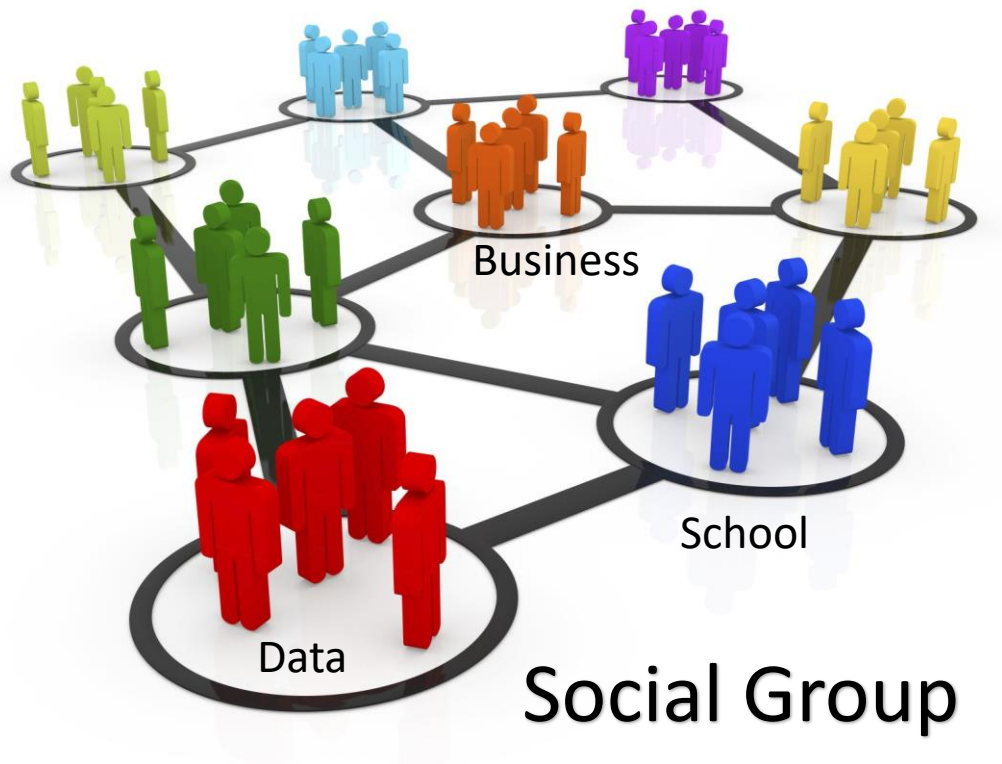


HW1

- ANOVA
- Sampling
- Regression
- Basic data analysis skills, visualization
- Dealing with real data



Collective info



Background

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

Group name	Group category	Group size	Message number	Message density	Sex ratio	Average year	Age gap	Regional concentration	Mobile ratio	Session number	Non-response ratio	Night-chat ratio	Picture ratio
**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
**秀嘉园	1	73	2	0.010274	0.57971	27.5	9.042524	0.40678	0	2	1	1	0
**^ 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
**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
**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
**踏联盟	1	60	22	0.038136	0.807692	29.7381	4.619109	0.357143	1	4	0.5	1	0.045455
**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
**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
**涂鸦天津+	1	63	48	0.043779	0.83871	23.93333	4.456705	0.716667	0.208333	18	0.722222	0.083333	0
ns万有引力--颓废1群	1	72	177	0.016041	0.873239	22.41538	4.495784	0.119403	0.175141	19	0.842105	0.327684	0.00565
**弟盟罗马战队总群	1	203	4615	0.019583	0.935323	28.2	5.07762	0.57754	0.044854	546	0.496337	0.300108	0.058072

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

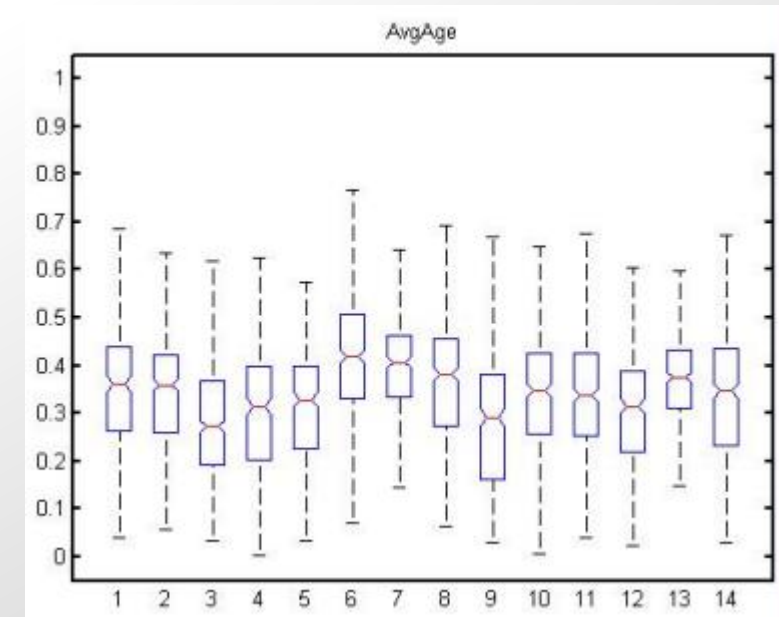
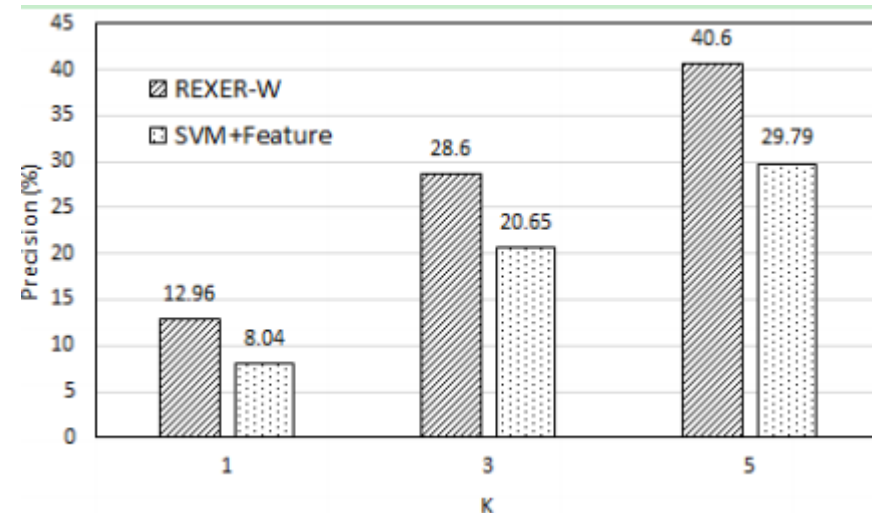
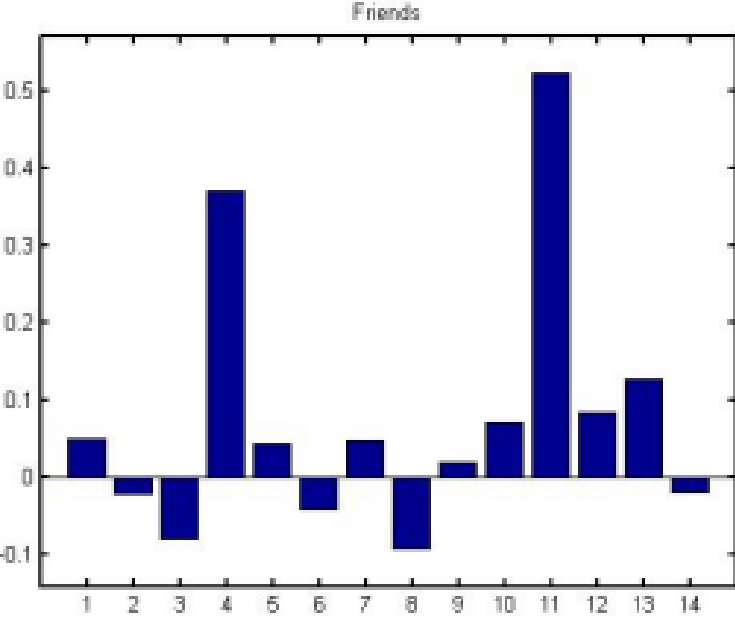
Feature can distinguish categories? => features varied significantly across the categories?

Data & Problem

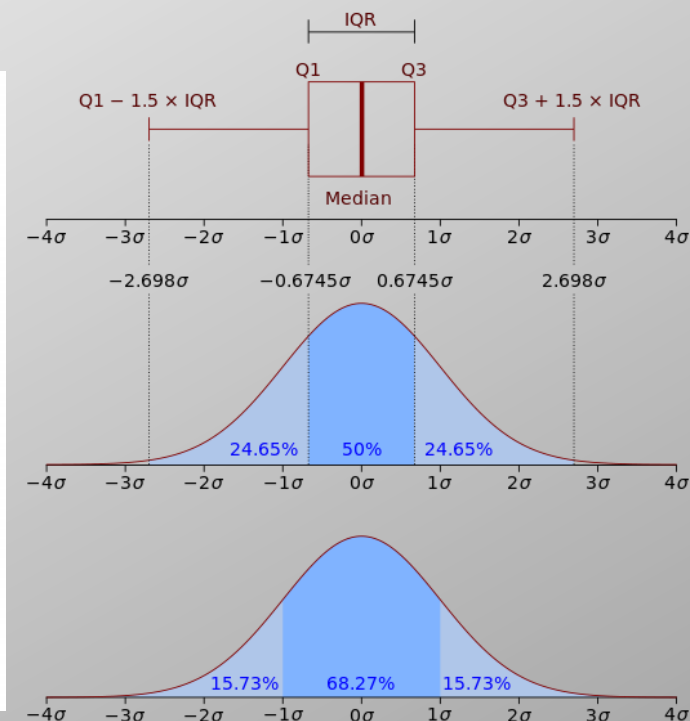
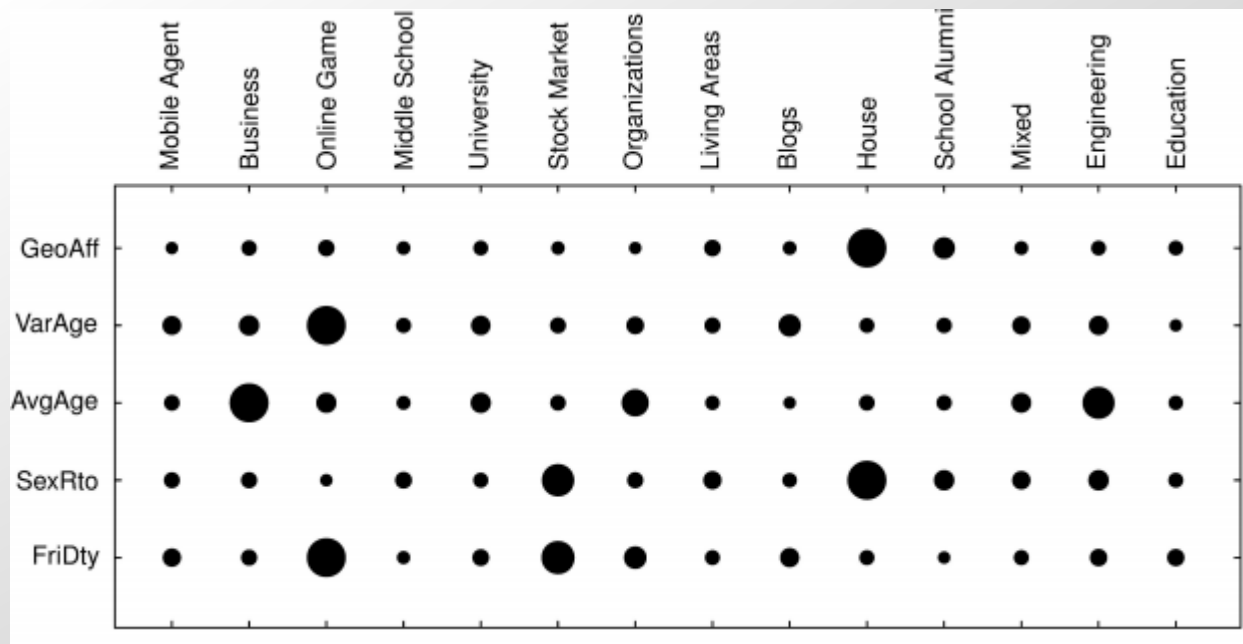
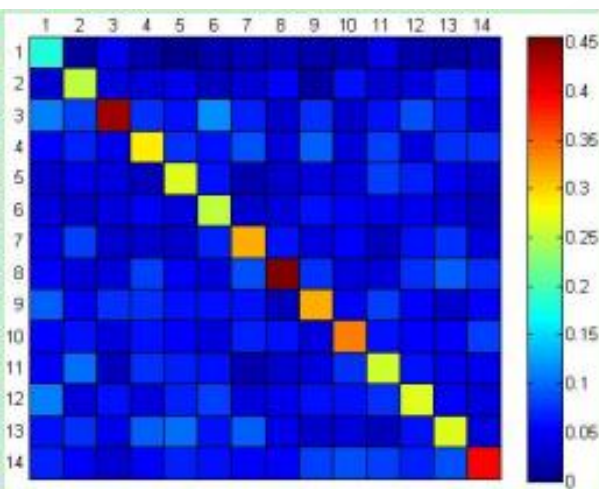
Does the chatting behaviors can be predicted by other features?

- Tools: Matlab, R, python, spss, excel...
- Detailed problem description, see homework1.pdf
 - Problem definition
 - Understand data and verify assumptions
 - Experiments
 - Visualization
- Hand in: an experiment report and codes
- What Not to do:
 - Plagiarism
 - Late submission

Requirements



2-D=>contain more information



Visualization