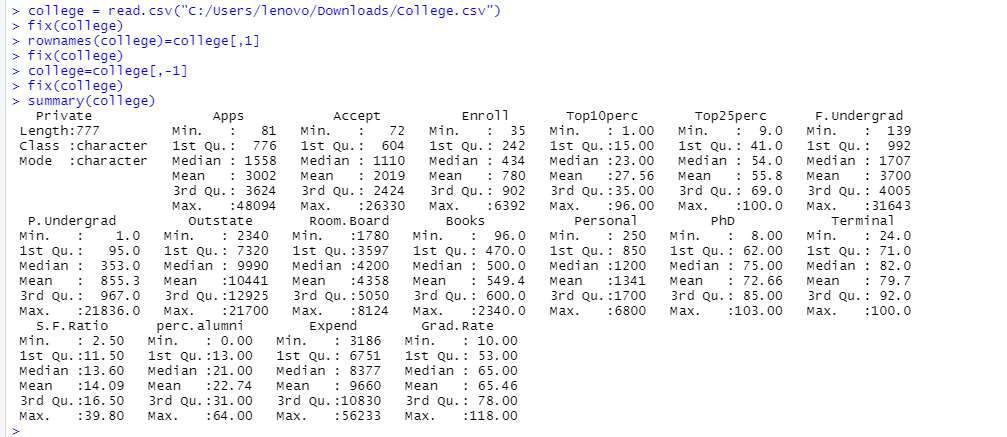
8.abc的代码截图



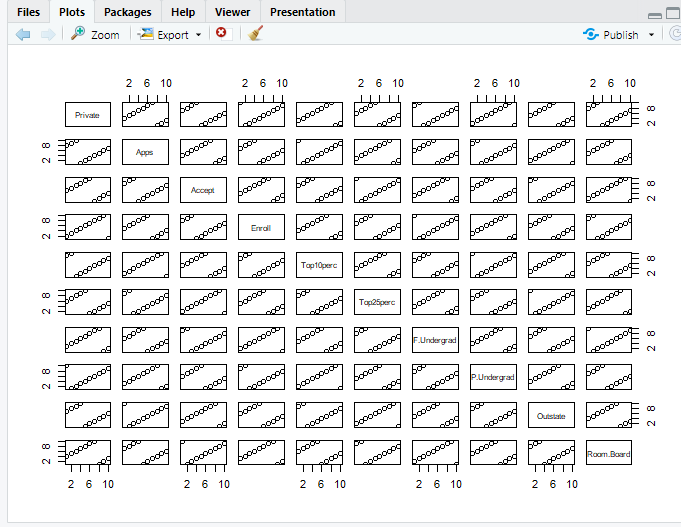
1. **read.csv()运行截图**



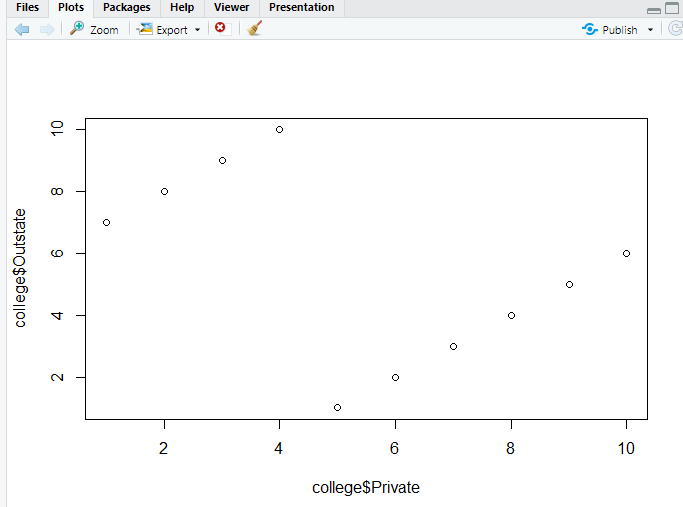
**b.用fix()函数观察数据（R为每行每列大学分配名字）**



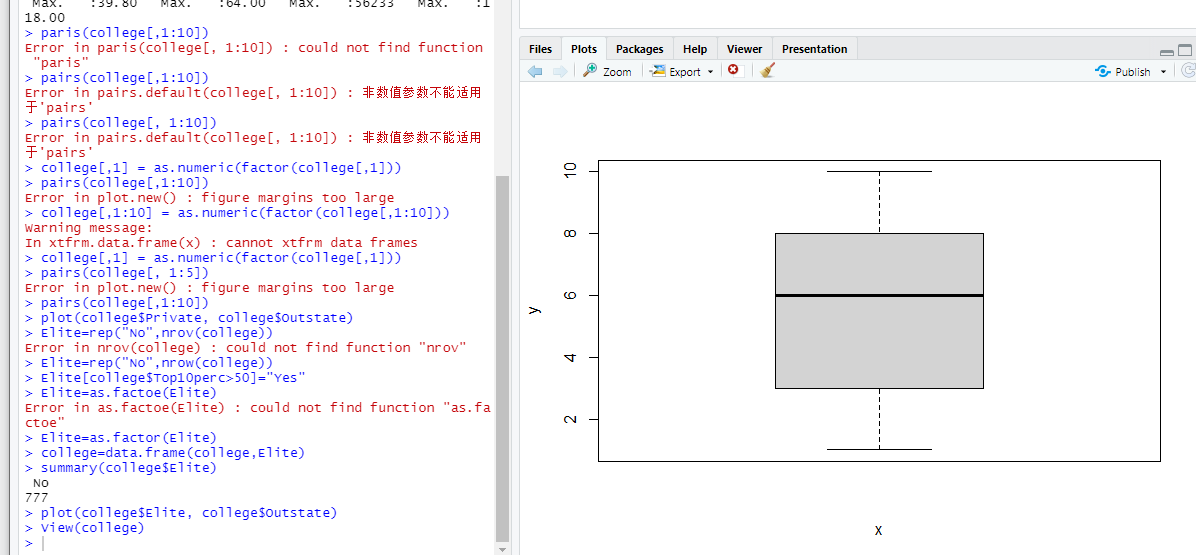
1. **i的summary（）函数截图和ii.pairs（）（汇总信息并对前十列或变量产生散点图矩阵）**



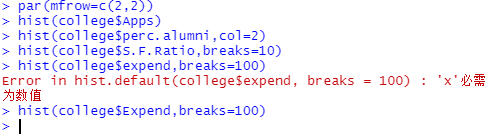
**c.iii用plot()函数产生Outstate对Private变量的沿边箱线图**

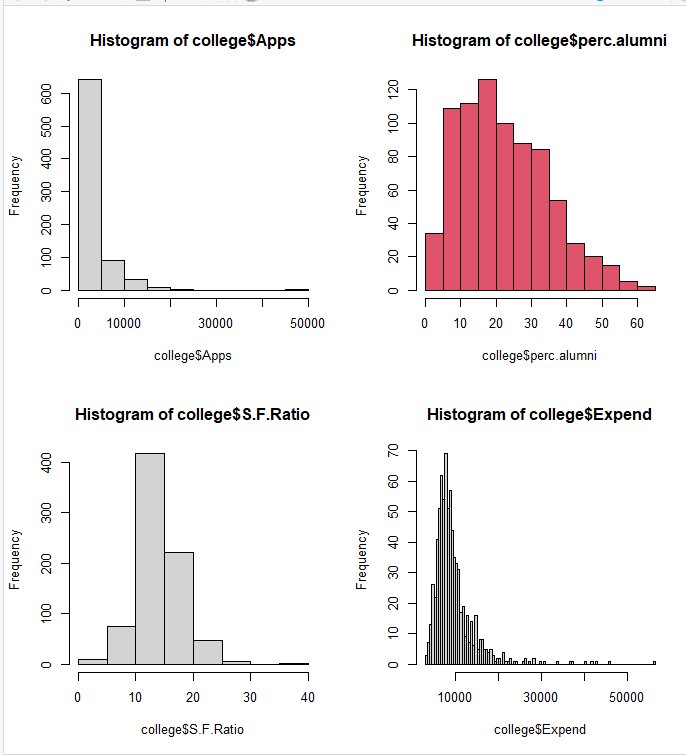


**c.iv的代码和运行截图**

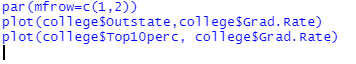


**c.v代码截图和运行截图**

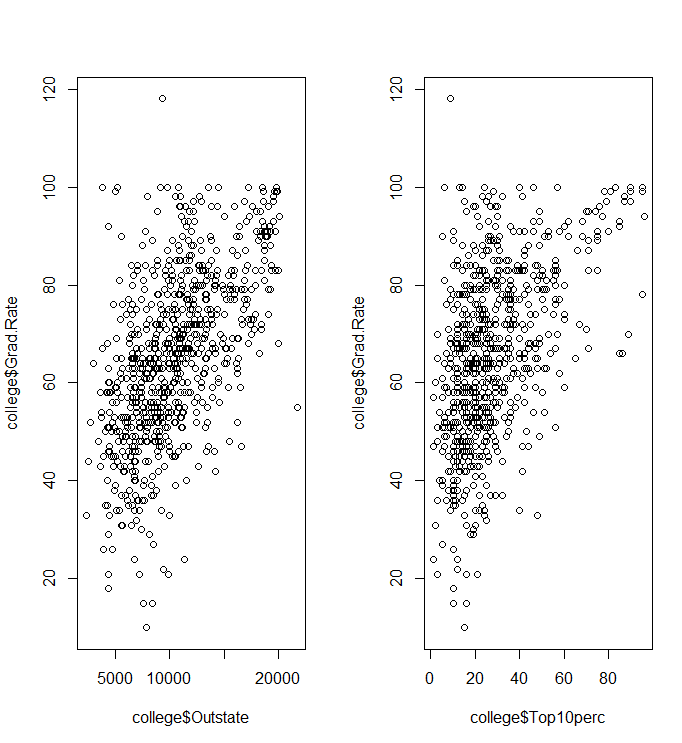




**c.vi**



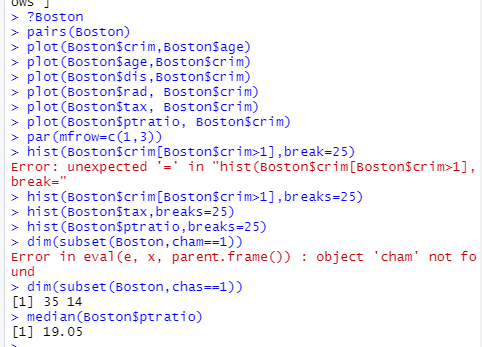
**运行截图**



**非本州学生学费越高，毕业率相对就越高。**

**从排名前10%的高中班毕业的新生的毕业率反而不是很高。**

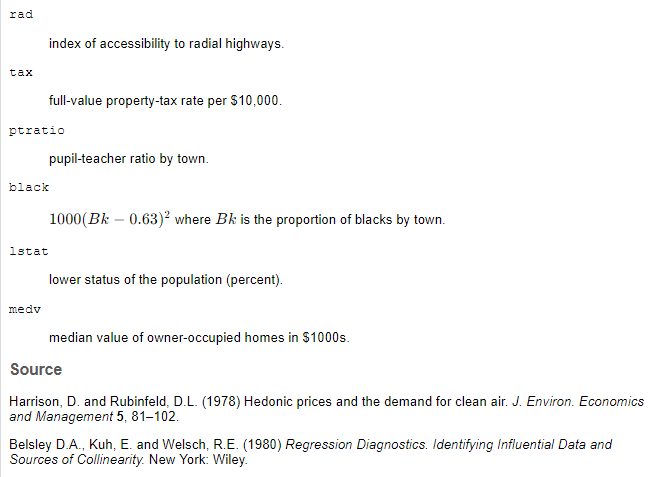
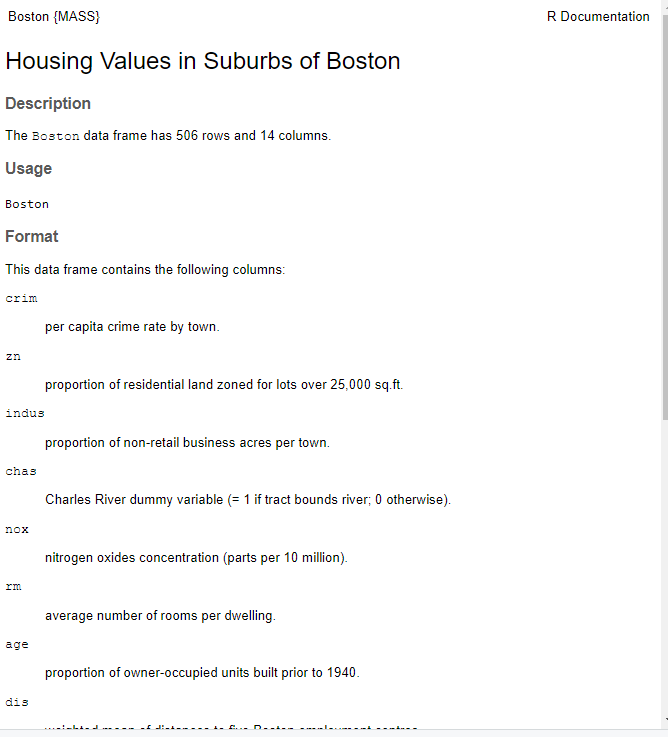
**10.（a）a-e的代码截图**



**>Library（MASS）**

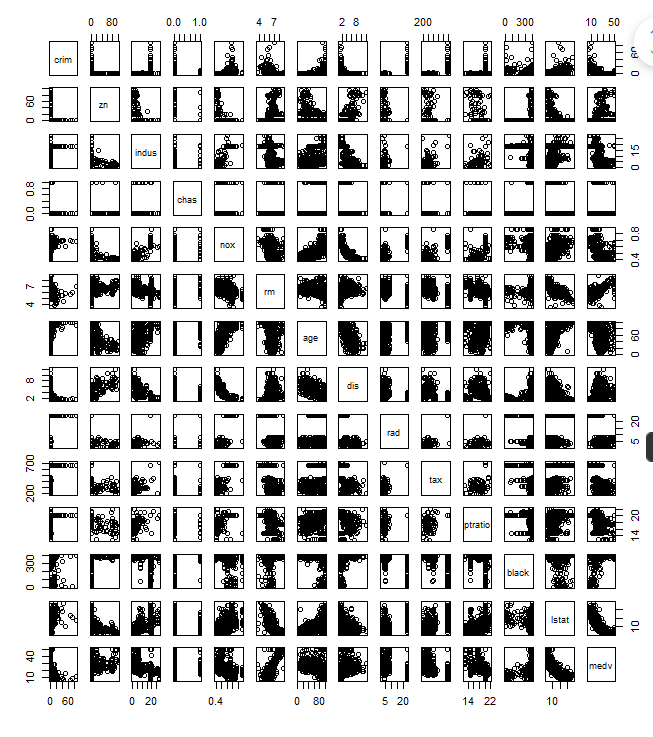
**>Boston**

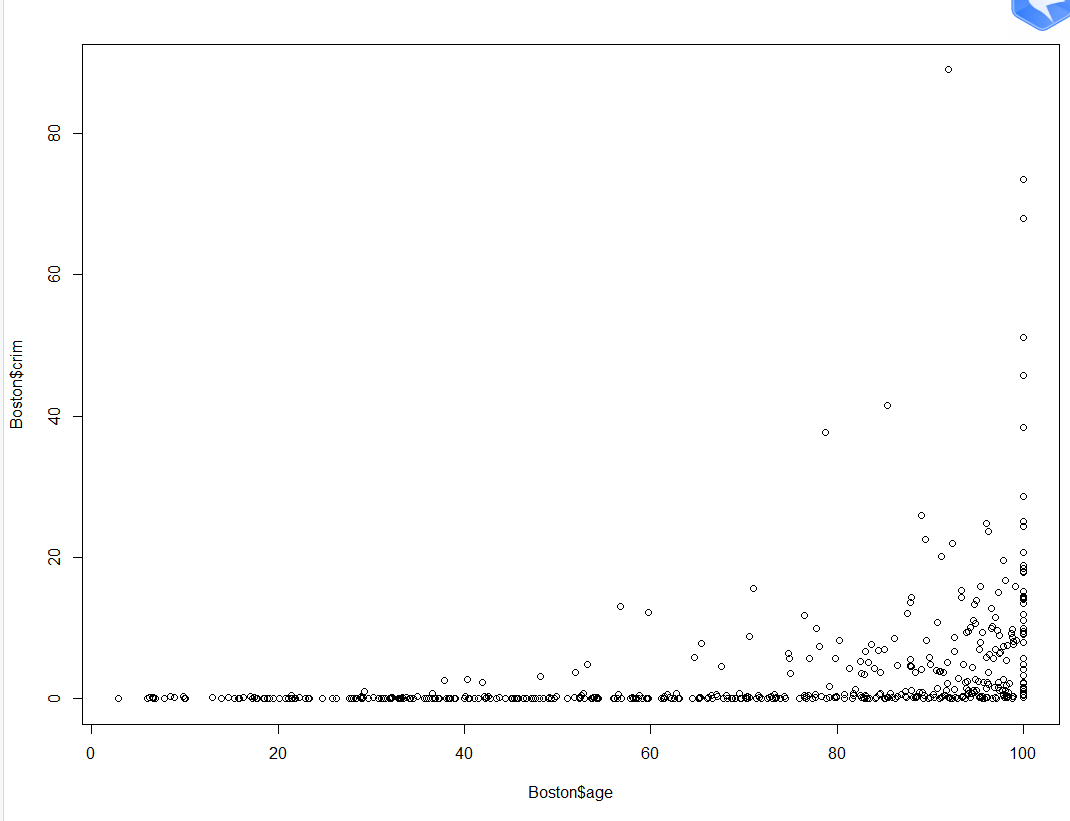
**>?Boston**

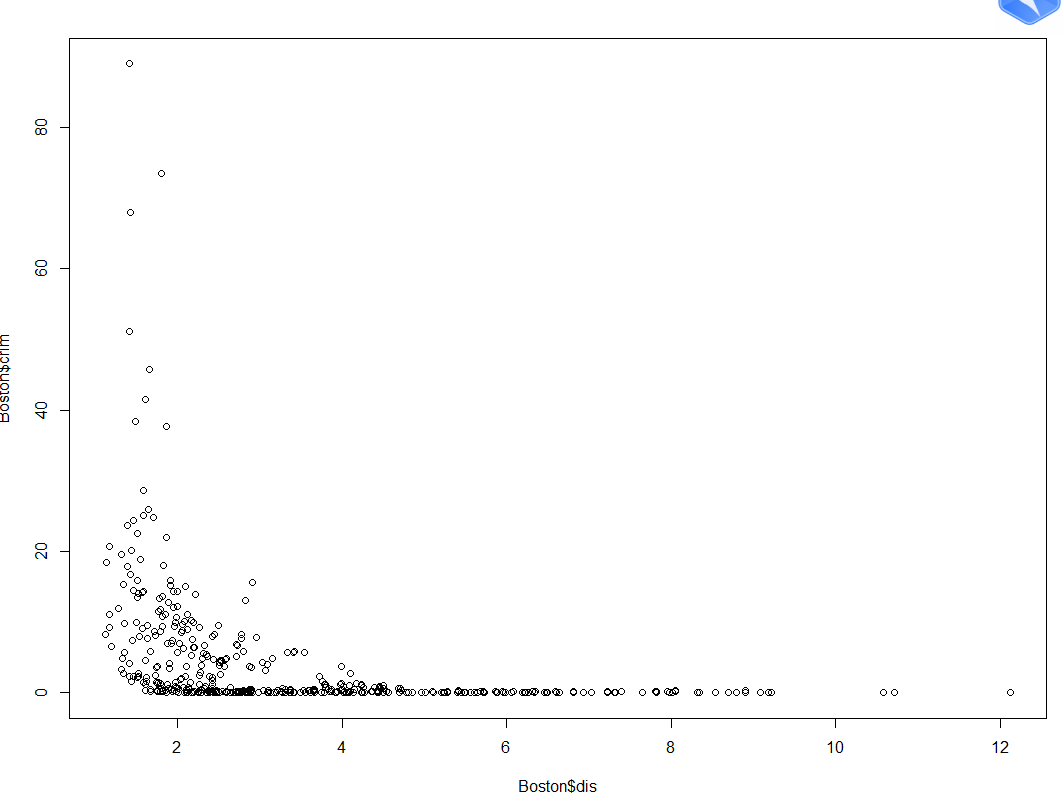


**可以看出这个数据有505行，14列。列代表每种数据如crim，zn，indus，chas等，每行就表示不同的具体数据。**

1. **散点图 发现一共有14\*14对关系，可以看到crim和age呈正比而和dis呈反比关系nox和dis呈反比关系等。crim 受其他变量影响变化明显。**

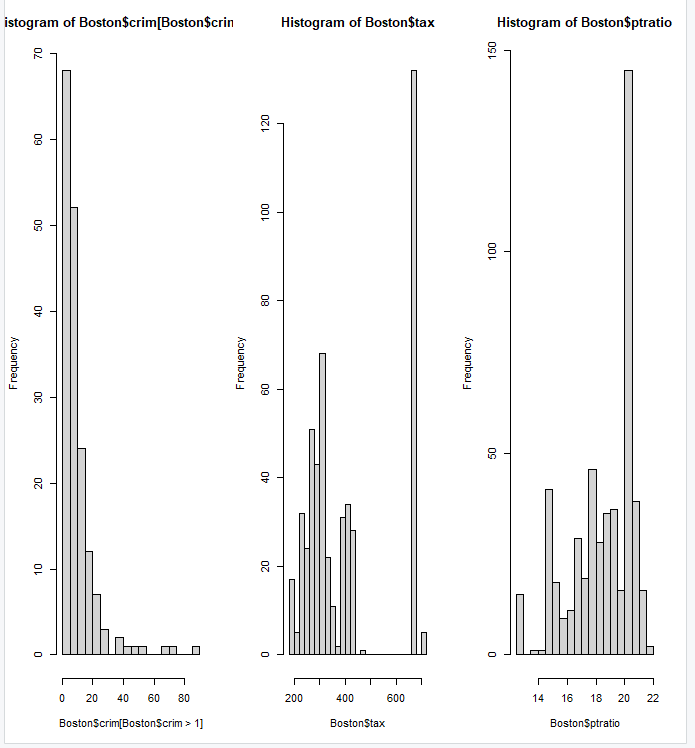


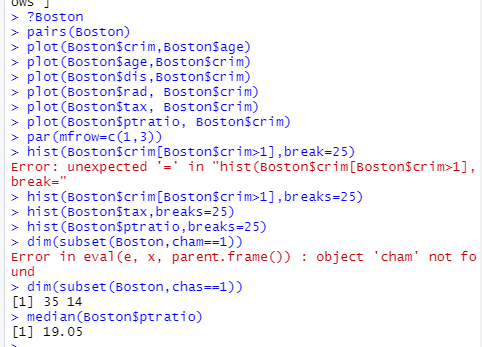
1. **发现crim和年龄有关，年龄偏高，犯罪率也会偏高。距离五个上班区域的加权平均距离dis 越低，高犯罪概率值越密集。**



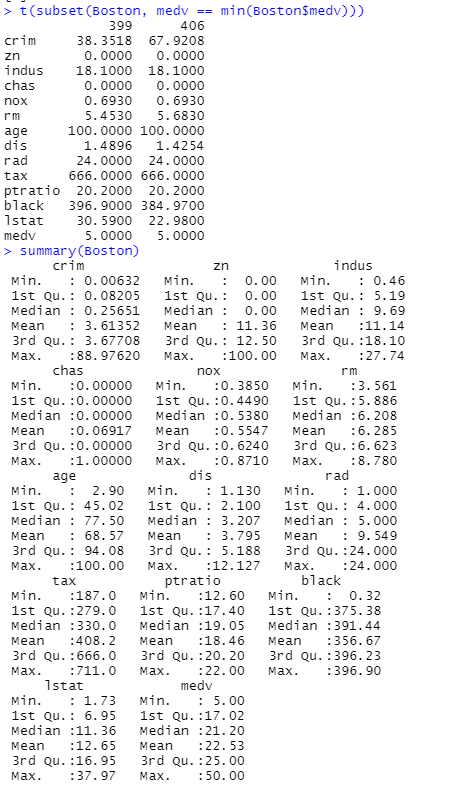
1. **截图**

**郊外的犯罪率不会特别高，但是税率还是蛮高的，超600的很多，师生比也比较高**



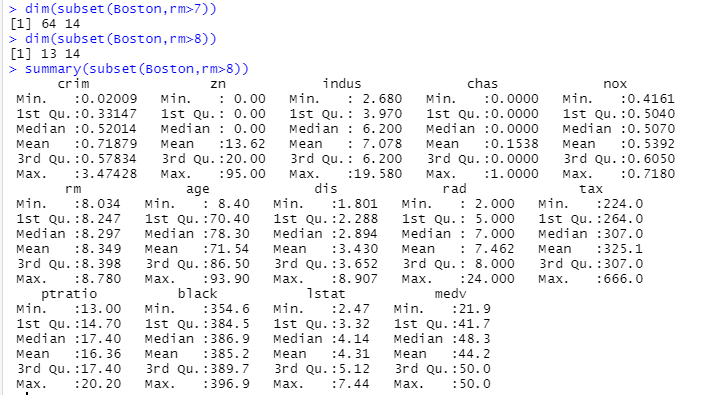
**e&f有35个郊区在查尔斯河岸附近，该数据集里城镇师生比的中位数为19.05**

**g.（1）业主自用住房的中位数最小的波士顿郊区是第399个。 其他预测变量的取值分别是38.35180,0，18.1,0,0.6930,5.4530,100.0，1.4896,24.000,666.000,20.2000,396.9000,30.5900，5.000. 在总体的分布上：犯罪率高，住宅用地比例低，零售商业比例高，不靠近河，氮氧化物浓度较高，住宅房间数不是很多，年龄很高，距离五个上班区域的加权平均距离较近， 交通发达，税率很高，师生比较高，黑人占比很高，**

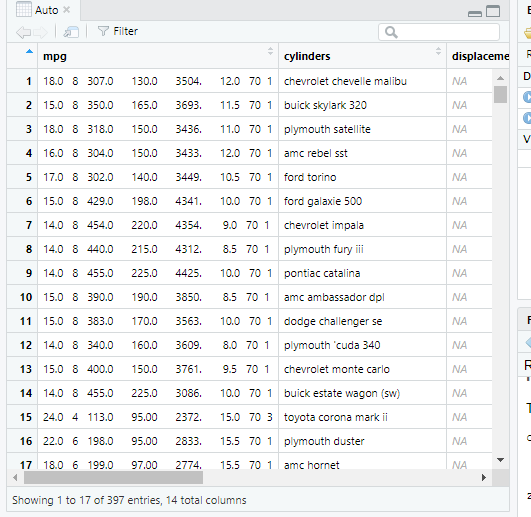
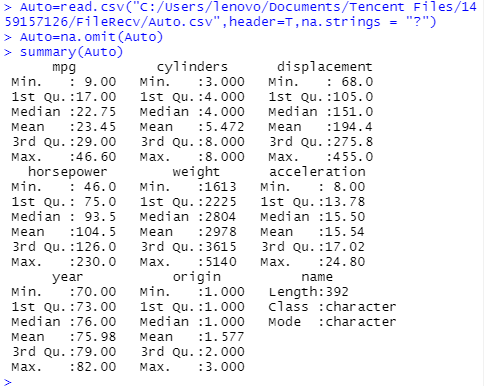


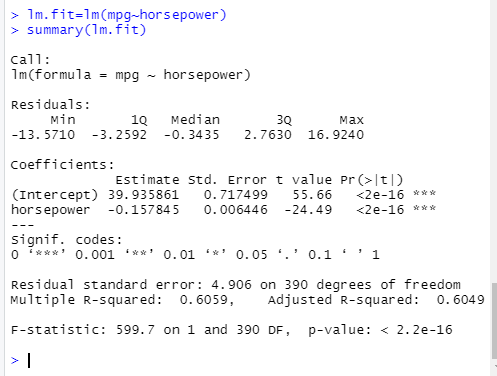
**h.有64个郊区居民平均居住房间数量超过7，有13个郊区数超过8个房间。**

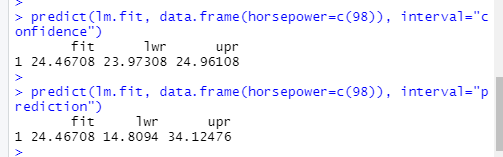
**居民平均居住房间数超过8个的郊区特征：犯罪率低，住宅用地比例高，零售商业比例较之前别的组的数据更合适更合理， 河，氮氧化物浓度处于中等水平，住宅房间数多，距离五个上班区域的加权平均距离近，交通发达，税率低，师生比例低，黑人占比极高，地位低的人群比例也处于较低状态**



**8.（a）打开auto**

****

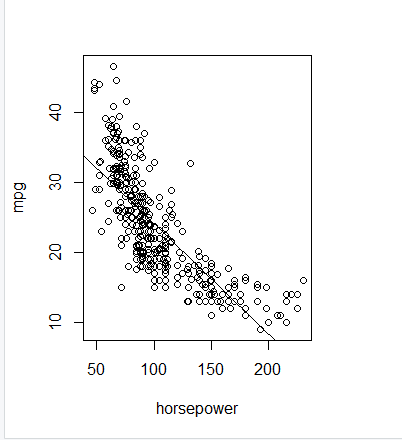
****

1. **由F值很大和p值很小可以看出，预测变量马力和响应变量油耗有关系。**
2. **由数据residual standard error和R-squared可知：lm的RSE拟合为4.906，R2 (lm)Fit约为0.6059，这意味着MPG中60.59%的方差是由马力解释的。**
3. **负相关。**
4. ****

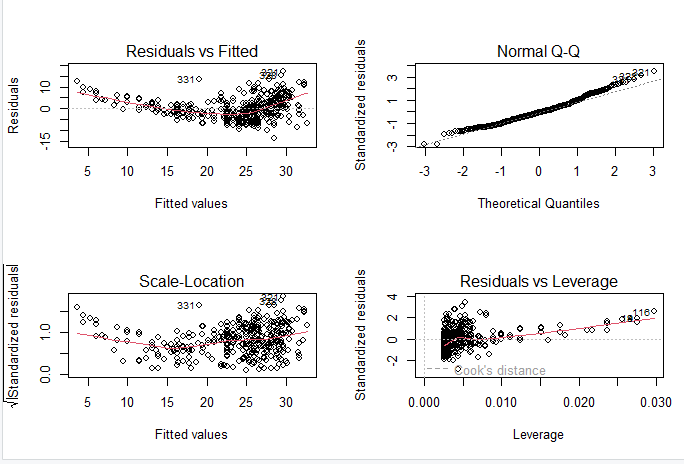
**置信区间【23.97308,24.96108】**

**预测区间【14.8094,34.12476】**

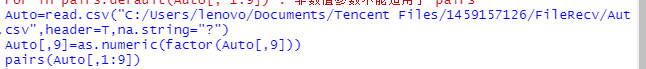
**8b.** ****

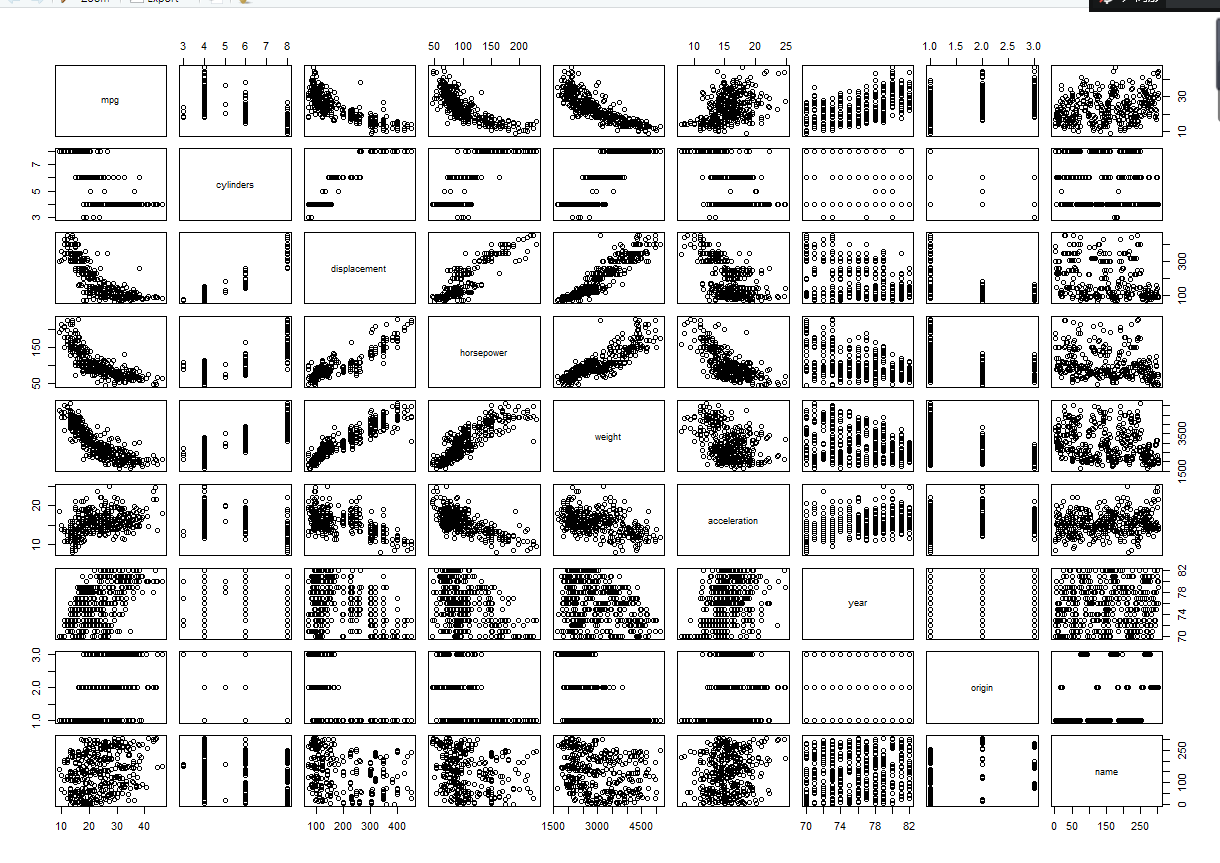
****

**8c**

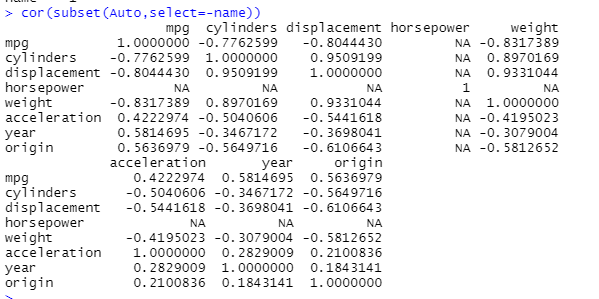
****

**9 (a)**

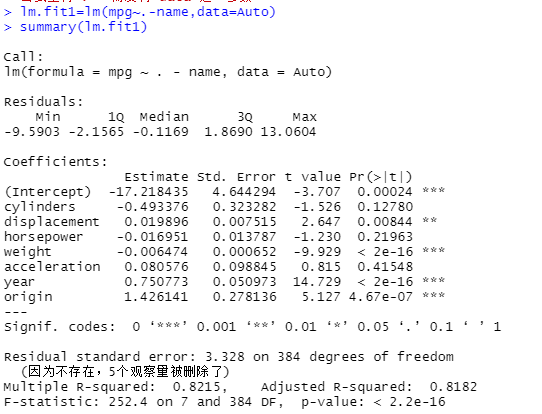
****

****

**9.(b)**

****

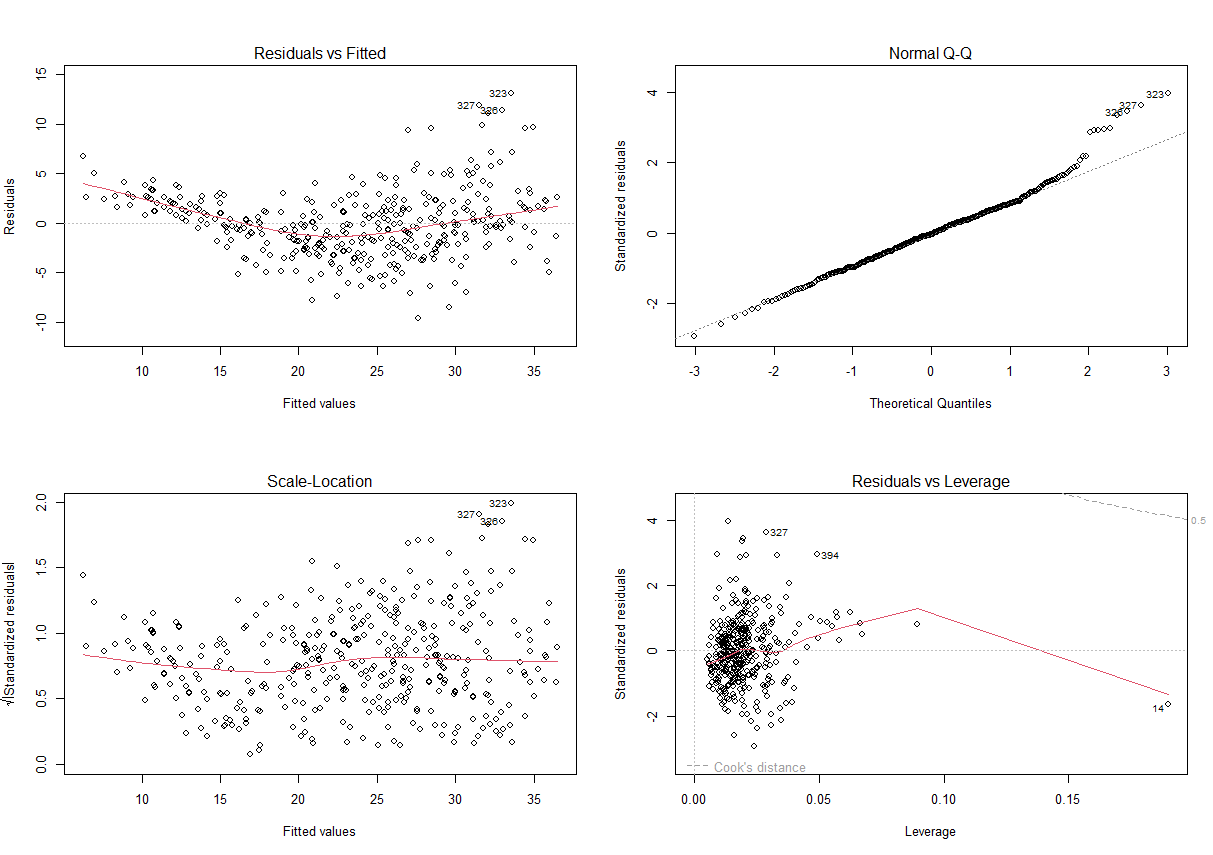
**9.c**

****

1. **有关系，因为F值远大于1且p值很小**
2. **Displacement，weight，year，origin与相应变量在统计上具有显著关系，因为由数据可以看出对应t值的p值很小说明显著性很强**
3. **Year和mpg的关系，即每增加一年，mpg增加0.75个单位**

**9（d）**

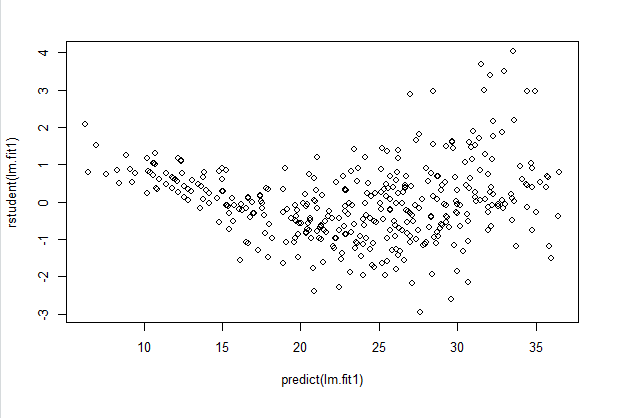
****

****

**图中点右侧有一些点有离群，但通过分析Normalqq图分析，基本呈正态分布，就是在结尾的部分有些偏离，但是不是非常异常。**

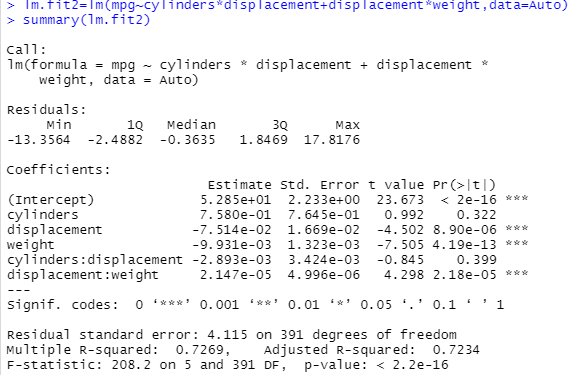
**杠杆图识别出点14是异常高杠杆作用的点。**

****

****

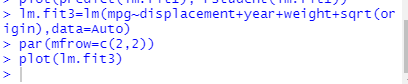
**有残差图中存在异常值，因为有大于3的值**

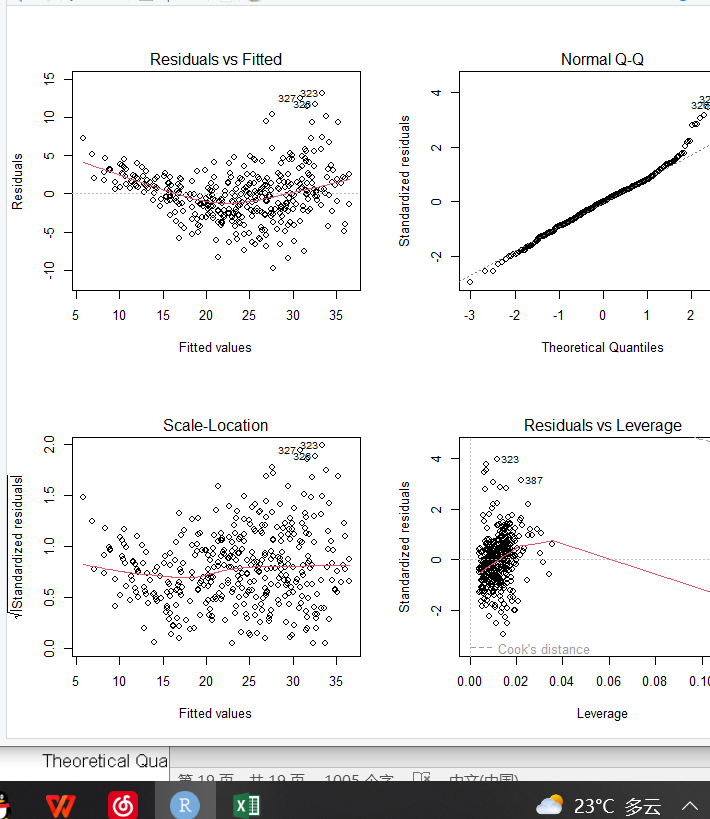
**9（e）**

****

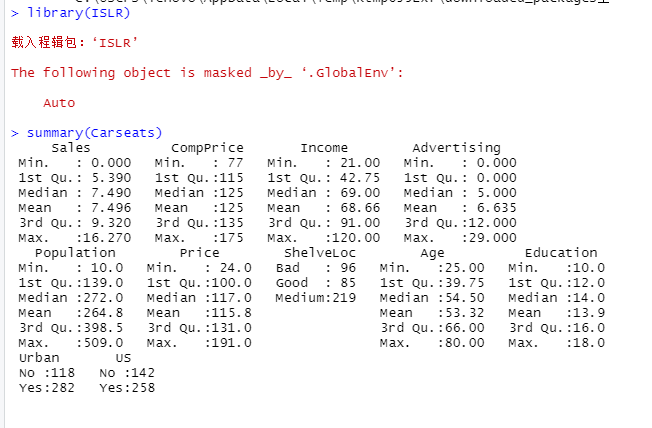
**有明显作用**

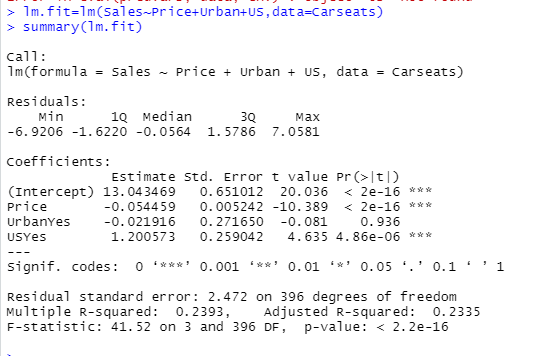
**9（f）**

****

****

**10（a）**

****

****

**10b**

Price：由给的数据可知，t统计量的p值较低，价格（Price）和销售额（Sales）之间存在关系。这个系数表明价格和销售额之间的负关系:当价格增加时，销售额减少。

Urbanyes：t统计值的高p值，商店的位置和销售额之间没有关系。

USyes：由t统计值的pr值，商店是否在美国与销售额之间存在关系。系数表明USYes和Sales之间呈正相关关系:如果商店在美国，销售额将增加

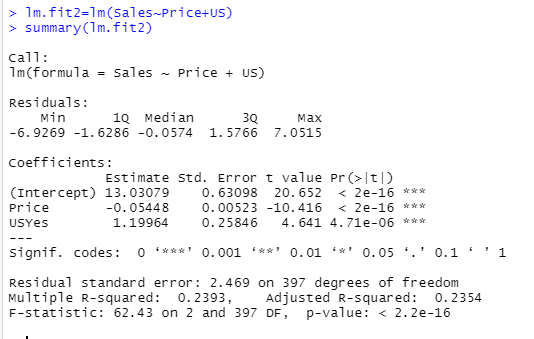
**10c**

**Sales=-0.054price-0.022UrbanYes+1.201USYes+13.043**

**10d**

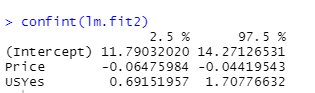
Price和USYes

**10e**



RSE在（a）中为2.472在（e）为2.469拟合的都很好，但是相比之下还是e中的小一些，所以e中的拟合更好一点

**10g**

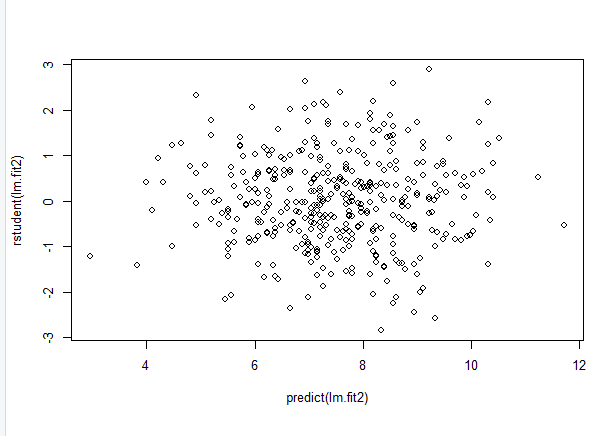
****

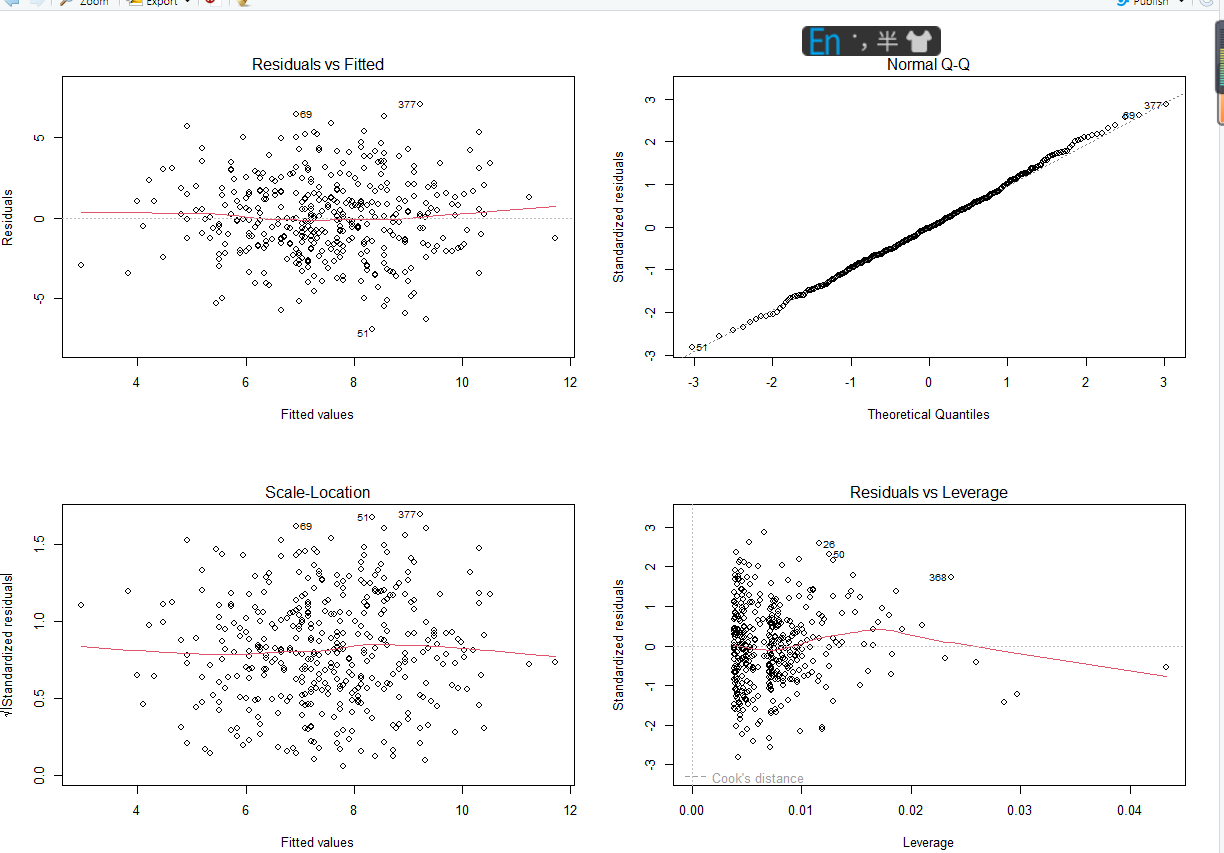
**95%的置信区间**

**【11.79032,14.27126531】**

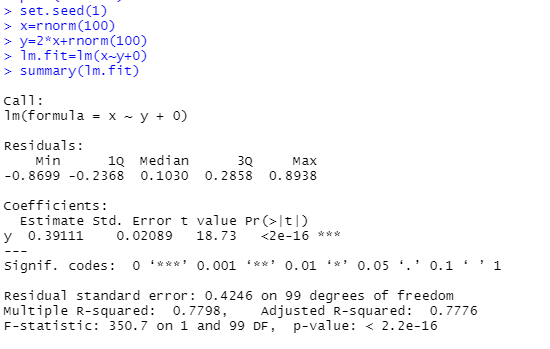
**10h**

****

****

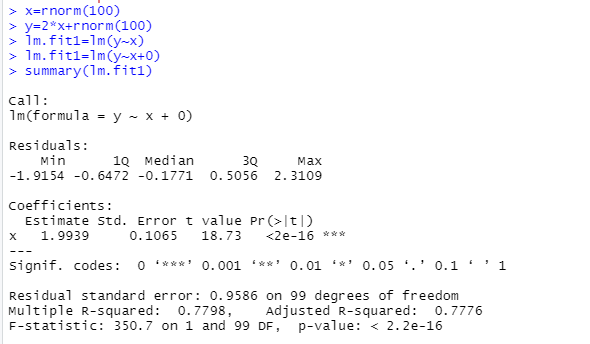
****

**11.a**

****

**估计系数：0.3911 标准差：0.02089，t统计量18.73，F值非常大，p值非常小，所以可以拒绝零假设。**

**11b**

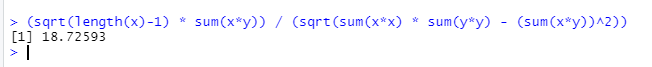
****

**估计系数：1.9939 标准差：0.1065，t统计量18.73，F值非常大，p值非常小，所以可以拒绝零假设。**

**11c**

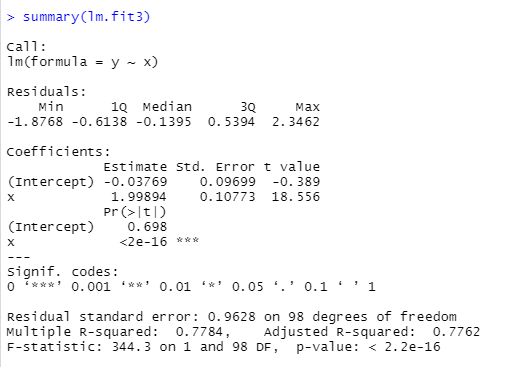
**可以写出线性回归方程 y=2x+ 或者也可以写成 x=1/2 \*（y-）**

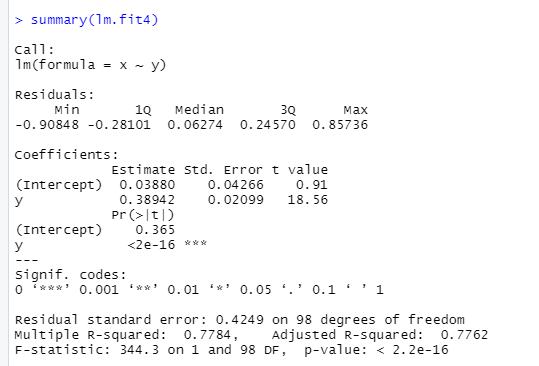
**11d**

****

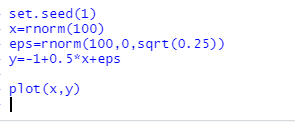
**11f**

****

****

****

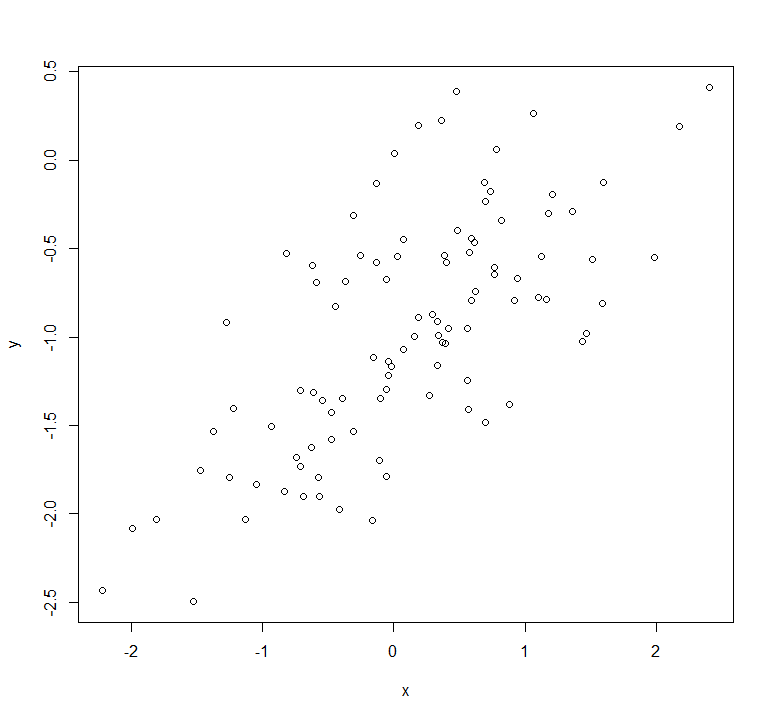
**13（a）-（d）**

****

**C：y的长度就是（-1+0.5\*x+eps）**

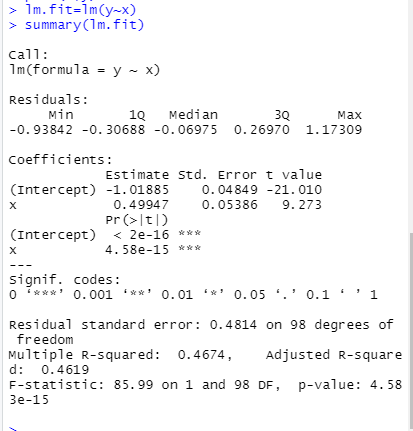
**β0 is -1, β1 is 0.5.**

**d**

****

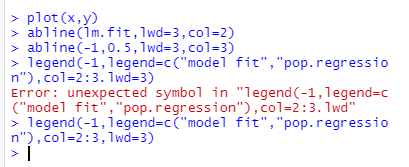
**可以看出x和y有线性回归关系，并且是正相关关系。**

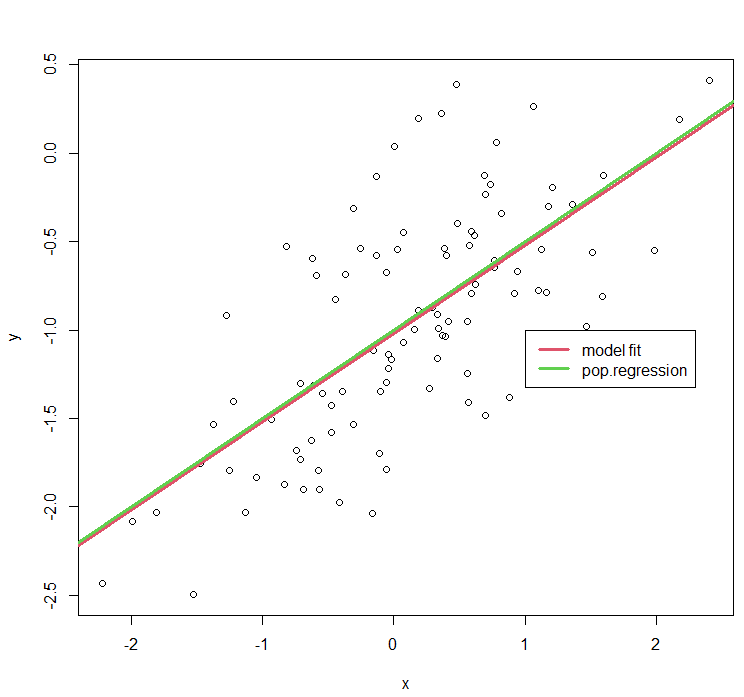
**13e**

****

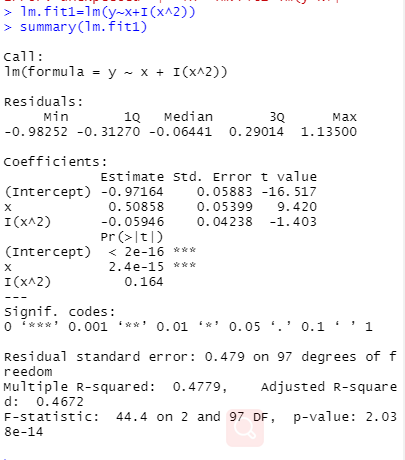
**真实系数值与估计值相近，同时由得到的数据可知F很大，p很小，可以拒绝H0（零假设），认为x和y有关系。**

**13（f）**

****

****

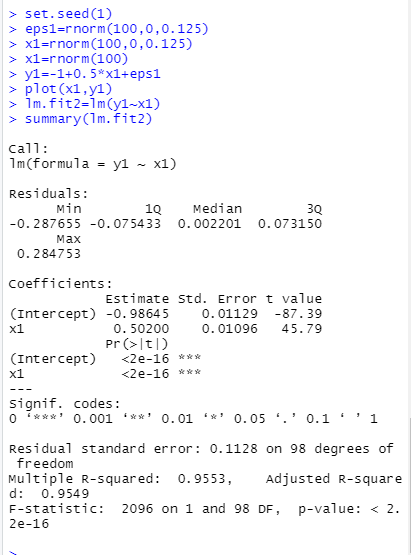
**13（g）**

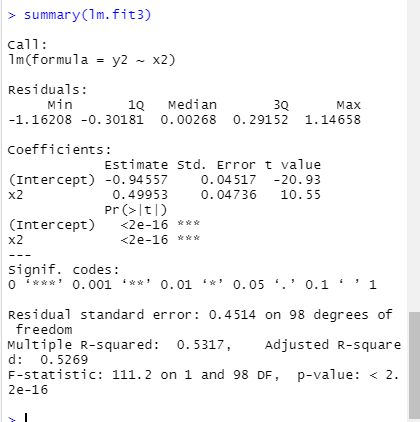
****

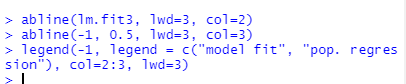
**RSE减小了，说明没有提高拟合度。**

**13g**

**13h**

****

****

****

**14（a）**