相关性分析的可视化_相关系数图的绘制过程

作者：王晓晓，赵一鸣 来源：临床流行病学和循证医学

本文原地址：https://www.iikx.com/news/statistics/1938.html

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

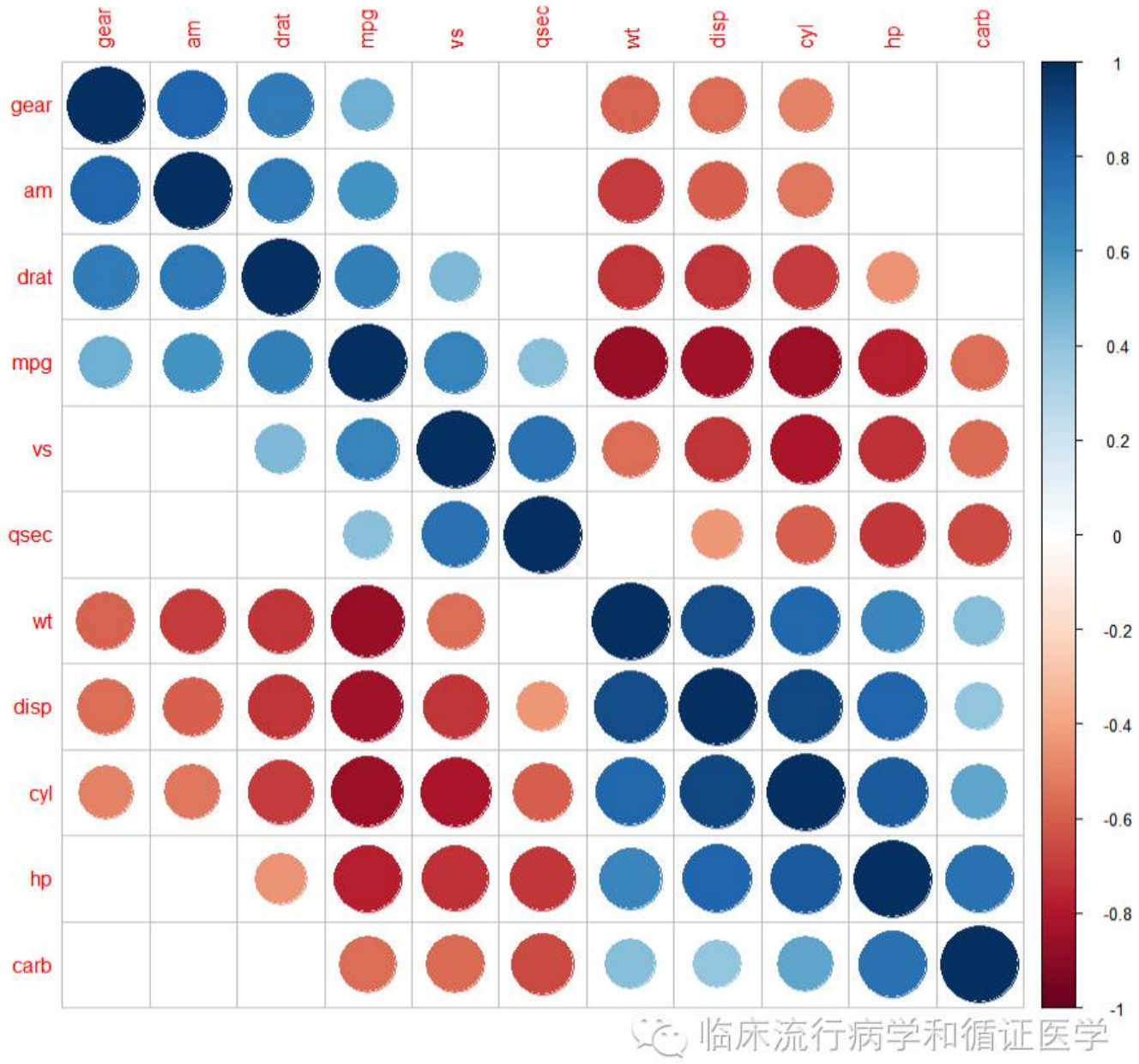
多个连续变量两两之间做简单相关，您还记得会输出什么样的结果吗?您喜欢这种输出方式吗?先放两张图，大家随意感受一下。

第一张，大家应该很熟悉，是11个变量两两之间的简单相关系数。这么多数字堆在一起，很难快速的发现变量之间的相关性，根本不想看第二眼，有没有?!

```
> M
             mpg        cyl       disp         hp        drat         wt
mpg   1.0000000 -0.8521620 -0.8475514 -0.7761684  0.68117191 -0.8676594
cyl  -0.8521620  1.0000000  0.9020329  0.8324475 -0.69993811  0.7824958
disp -0.8475514  0.9020329  1.0000000  0.7909486 -0.71021393  0.8879799
hp   -0.7761684  0.8324475  0.7909486  1.0000000 -0.44875912  0.6587479
drat  0.6811719 -0.6999381 -0.7102139 -0.4487591  1.00000000 -0.7124406
wt   -0.8676594  0.7824958  0.8879799  0.6587479 -0.71244065  1.0000000
qsec  0.4186840 -0.5912421 -0.4336979 -0.7082234  0.09120476 -0.1747159
vs    0.6640389 -0.8108118 -0.7104159 -0.7230967  0.44027846 -0.5549157
am    0.5998324 -0.5226070 -0.5912270 -0.2432043  0.71271113 -0.6924953
gear  0.4802848 -0.4926866 -0.5555692 -0.1257043  0.69961013 -0.5832870
carb -0.5509251  0.5269883  0.3949769  0.7498125 -0.09078980  0.4276059
             qsec         vs          am       gear        carb
mpg   0.41868403  0.6640389  0.59983243  0.4802848 -0.55092507
cyl  -0.59124207 -0.8108118 -0.52260705 -0.4926866  0.52698829
disp -0.43369788 -0.7104159 -0.59122704 -0.5555692  0.39497686
hp   -0.70822339 -0.7230967 -0.24320426 -0.1257043  0.74981247
drat  0.09120476  0.4402785  0.71271113  0.6996101 -0.09078980
wt   -0.17471588 -0.5549157 -0.69249526 -0.5832870  0.42760594
qsec  1.00000000  0.7445354 -0.22986086 -0.2126822 -0.65624923
vs    0.74453544  1.0000000  0.16834512  0.2060233 -0.56960714
am   -0.22986086  0.1683451  1.00000000  0.7940588  0.05753435
gear -0.21268223  0.2060233  0.79405876  1.0000000  0.27407284
carb -0.65624923 -0.5696071  0.05753435  0.2740728  1.00000000
>
>
>
```

临床流行病学和循证医学

大家再来看第二张，是不是会好很多。除了颜值，咱们还得看看它究竟能提供一些什么信息。



首先，左边界和上边界显示的是11个变量名。其次，不同颜色填充的实心圆代表了相应的相关系数。大家看右侧的颜色条，蓝色系的为正相关，红色系的为负相关，越靠近颜色条两头，相关系数越大。所以说，我们可以通过实心圆的颜色和大小判断相关的方向和相关系数的大小。

另外，还有一些是没有实心圆的，这是为什么呢?因为我们把p>0.05的隐藏了，也就是说现在呈现的是p<0.05的变量间的相关。说这么多，这个看起来还不错的图怎么画呢?

今天咱们以R自带的数据为例，一起来重现下
这张**相关系数图**
(英文叫corrplot)。数据要求相同研究对象的数据在同一行，相同变量的数据在同一列，可存储为csv格式，利用read.csv(file.choose())读取数据。

安装R和RStudio软件的过程咱就不说了，直接到RStudio界面。

1. 安装程序包，install.packages(corrplot)

2. 加载程序包，library(corrplot)

3. 调用R自带的数据，data(mtcars)

4. 计算mtcars的相关系数矩阵，M<-cor(mtcars)

5接下来的16行都是为了编写函数cor.mtest，获取相关系数的p值。不过大家勿担心，直接copy就可以了，不需要更改任何参数。

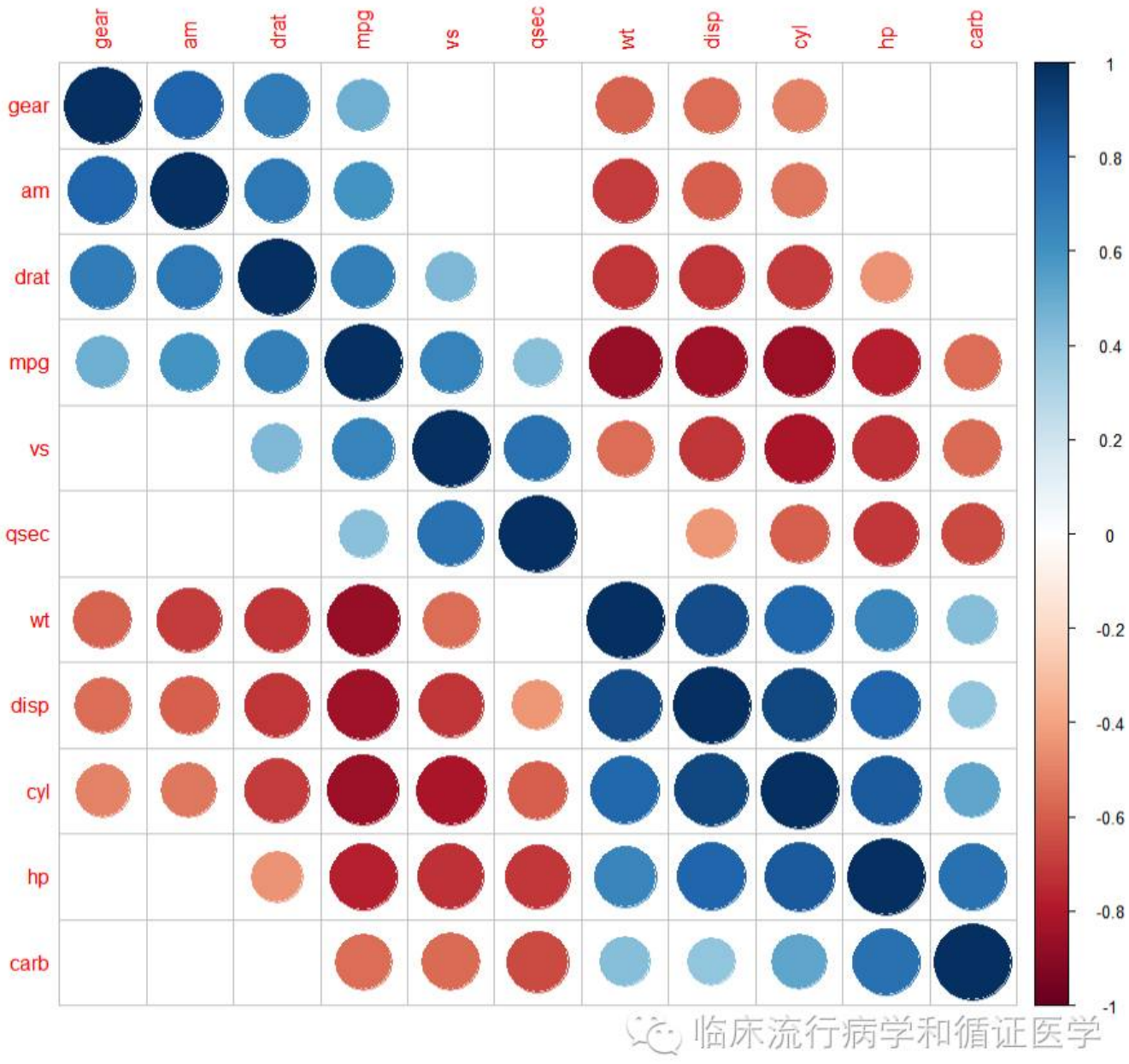
```
1cor.mtest<- function(mat, conf.level = 0.95){
2mat <- as.matrix(mat)
3n <- ncol(mat)
4p.mat <- lowCI.mat <- uppCI.mat <-matrix(NA, n, n)
5diag(p.mat) <- 0
6diag(lowCI.mat) <- diag(uppCI.mat) <- 1
7for(i in 1:(n-1)){
8for(j in (i+1):n){
9tmp <- cor.test(mat[,i], mat[,j],conf.level = conf.level)
10p.mat[i,j] <- p.mat[j,i] <-tmp$p.value
11lowCI.mat[i,j] <- lowCI.mat[j,i] <-tmp$conf.int[1]
12uppCI.mat[i,j] <- uppCI.mat[j,i] <-tmp$conf.int[2]
13 }
14 }
15return(list(p.mat, lowCI.mat, uppCI.mat))
16}
```

6. 利用函数cor.mtest，得到p值，res <- cor.mtest(mtcars,0.95)

7.绘制相关分析图

corrplot(M,order="AOE",p.mat=res[[1]],sig.level=0.05,insig ="blank")

其中M为相关分析系数矩阵，order="AOE"是对变量进行排序，p.mat=res[[1]]是为了获取相关分析p值，sig.level=0.05，insig = "blank"，主要是让p>0.05的不显示其相关系数。



运行上述程序后，右下角就是我们想要的图了。欢迎大家留言讨论。

更多 统计方法 请访问 https://www.iikx.com/news/statistics/

本文版权归原作者所有，请勿用于商业用途，爱科学iikx.com转发