dplyr\_join

```{r} #下面内容来自互联网，为了方便学习dplyr包里的join合并数据操作 #主要数据和内容参考：<https://stat545-ubc.github.io/bit001_dplyr-cheatsheet.html>

# 加载需要的包

library(dplyr) library(plyr) library(stringr) library(knitr)

# 定义一个函数来处理将array转换为data.frame的列命名问题，将第一行变为列名，然后去掉第一行生成df

norm\_df<-function(df){ colnames(df)<-df[1,] #定义上列名 df<-as.data.frame(df[-1,]) %>% #去掉首行 tbl\_df() #生成dplyr的tbl\_df对象  
 return(df) }

# 原始数据链式处理---------------------------------------

superheroes <- c(" name, alignment, gender, publisher", " Magneto, bad, male, Marvel", " Storm, good, female, Marvel", "Mystique, bad, female, Marvel", " Batman, good, male, DC", " Joker, bad, male, DC", "Catwoman, bad, female, DC", " Hellboy, good, male, Dark Horse Comics") %>% laply(strsplit,",") %>% #把vector根据，拆分成array aaply(1,str\_trim) %>% #去掉每个character元素里的首尾空格，基于stringr包 norm\_df()

publishers <- c("publisher, yr\_founded", " DC, 1934", " Marvel, 1939", " Image, 1992") %>% laply(strsplit,",") %>% #把vector根据，拆分成array aaply(1,str\_trim) %>% #去掉每个character元素里的首尾空格 norm\_df()

# 演示inner\_join----------------------------

ijsp <- inner\_join(superheroes, publishers) ijsp #Hellboy观测在合并数据中消失了，因为他的publisher是Dark Horse Comics并未在y的publisher中出现 #因为inner join是只显示匹配的（两边都有的）观测

# 演示semi\_join-----------------------------

sjsp <- semi\_join(superheroes, publishers) sjsp #结果与inner\_join类似，但只得到了有左边的变量

```