

# talk06 练习与作业

## 目录

0.1 练习和作业说明 . . . . .	1
0.2 Talk06 及 talk06-practices 内容回顾 . . . . .	1
0.3 练习与作业：用户验证 . . . . .	2
0.4 练习与作业 1: tidyr . . . . .	2
0.5 练习与作业 2: 作图 . . . . .	3
0.6 练习与作业 3: 数据分析 . . . . .	6

### 0.1 练习和作业说明

将相关代码填写入以 “{r}” 标志的代码框中，运行并看到正确的结果；

完成后，用工具栏里的”Knit” 按键生成 PDF 文档；

将 PDF 文档改为：姓名-学号-talk06 作业.pdf，并提交到老师指定的平台/钉群。

### 0.2 Talk06 及 talk06-practices 内容回顾

1. tidyr
2. 3 个生信任务的 R 解决方案
3. forcats

### 0.3 练习与作业：用户验证

请运行以下命令，验证你的用户名。

如你当前用户名不能体现你的真实姓名，请改为拼音后再运行本作业！

```
Sys.info()[["user"]]
```

```
## [1] "wchen"
```

```
Sys.getenv("HOME")
```

```
## [1] "/Users/wchen"
```

### 0.4 练习与作业 1: tidyr

---

#### 0.4.1 使用 grades 变量做练习

1. 装入 grades 变量；

```
library(dplyr);
```

```
grades <- read_tsv( file = "data/talk05/grades.txt" );
```

2. 使用 tidyr 包里的 pivot\_longer 和 pivot\_wider 函数对 grades 变量进行宽长转换；

```
## 代码写这里，并运行；
```

3. 使用 pivot\_longer 时，有时会产生 na 值，如何使用此函数的参数去除带 na 的行？

```
## 代码写这里，并运行；
```

4. 以下代码有什么作用？

```
grades %>% complete( name, course )
```

答:

## 0.5 练习与作业 2: 作图

---

### 0.5.1 用下面的数据作图

1. 利用下面代码读取一个样本的宏基因组相对丰度数据

```
abu <-  
  read_delim(  
    file = "../data/talk06/relative_abundance_for_RUN_ERR1072629_taxonlevel_species.txt",  
    delim = "\t", quote = "", comment = "#");
```

2. 取前 5 个丰度最高的菌，将其它的相对丰度相加并归为一类 Qita;
3. 用得到的数据画如下的空心 pie chart:

```
## 代码写这里，并运行;
```

---

### 0.5.2 使用 starwars 变量做图

1. 统计 starwars 中 hair\_color 的种类与人数时，可用下面的代码:

但是，怎么做到按数量从小到大排序?

```
library(dplyr)
```

```
##
```

```
## Attaching package: 'dplyr'
```

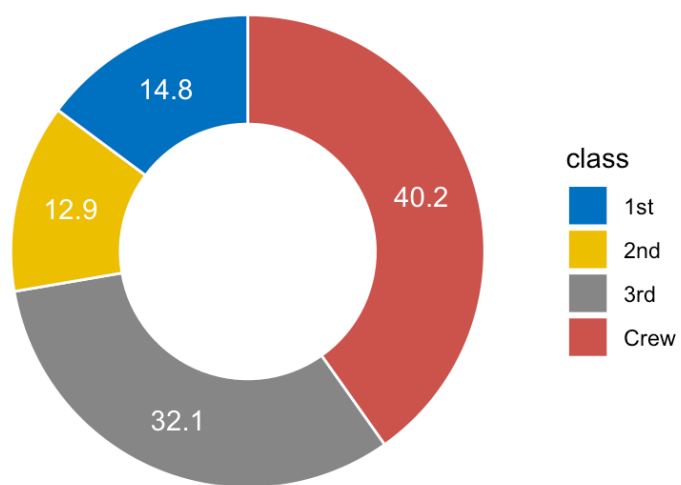
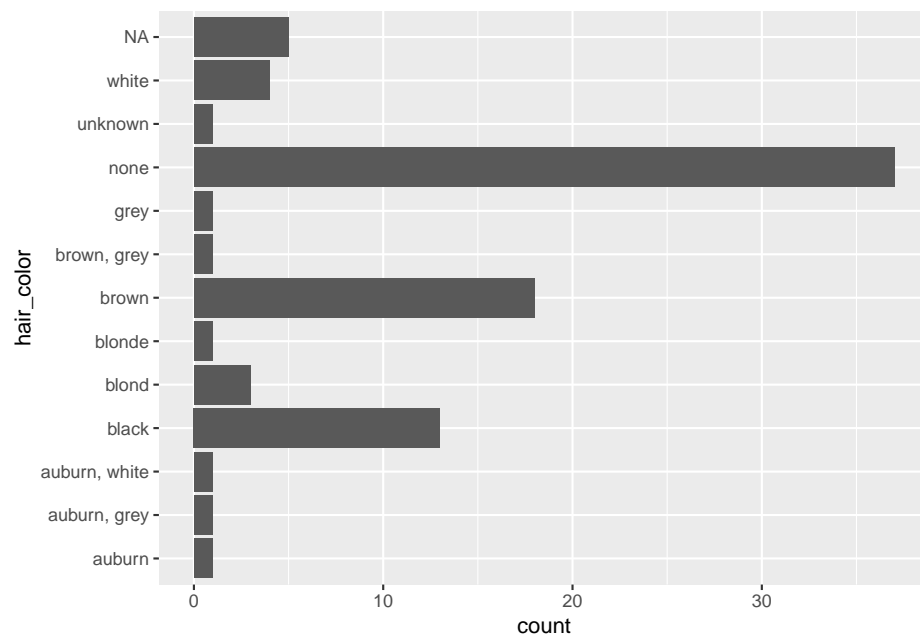


图 1: make a pie chart like this using the metagenomics data

```
## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
library(ggplot2)
library(forcats)
ggplot(starwars, aes(x = hair_color)) +
  geom_bar() +
  coord_flip()
```



```
## 代码写这里，并运行；
```

2. 统计 `skin_color` 时，将出现频率小于 0.05（即 5%）的颜色归为一类 `Others`，按出现次数排序后，做与上面类似的 barplot；

```
## 代码写这里，并运行；
```

3. 使用 2 的统计结果，但画图时，调整 bar 的顺序，使得 Others 处于第 4 的位置上。提示，可使用 `fct_relevel` 函数；

```
## 代码写这里，并运行；
```

## 0.6 练习与作业 3：数据分析

---

### 0.6.1 使用 STRING PPI 数据分析并作图

1. 使用以下代码，装入 PPI 数据；

```
ppi <- read_delim( file = "../data/talk06/ppi900.txt.gz", col_names = T,  
                  delim = "\t", quote = "" );
```

2. 随机挑选一个基因，得到类似于本章第一部分的互作网络图；

```
## 代码写这里，并运行；
```

### 0.6.2 对宏基因组相对丰度数据进行分析

1.data/talk06 目录下有 6 个文本文件，每个包含了一个宏基因组样本的分析结果：

```
relative_abundance_for_curated_sample_PRJEB6070-DE-073_at_taxonlevel_species.txt  
relative_abundance_for_curated_sample_PRJEB6070-DE-074_at_taxonlevel_species.txt  
relative_abundance_for_curated_sample_PRJEB6070-DE-075_at_taxonlevel_species.txt  
relative_abundance_for_curated_sample_PRJEB6070-DE-076_at_taxonlevel_species.txt  
relative_abundance_for_curated_sample_PRJEB6070-DE-077_at_taxonlevel_species.txt
```

2. 分别读取以上文件，提取 `scientific_name` 和 `relative_abundance` 两列；
3. 添加一列为样本名，比如 PRJEB6070-DE-073, PRJEB6070-DE-074 ... ；

4. 以 `scientific_name` 为 `key`，将其内容合并为一个 `data.frame` 或 `tibble`，其中每行为一个样本，每列为样本的物种相对丰度。注意：用 `join` 或者 `spread` 都可以，只要能解决问题。
5. 将 `NA` 值改为 0。

```
## 代码写这里，并运行；
```