

## Wrangle report

这个项目数据清洗的过程总共包括三个步骤：

第一步是收集，将三个不同格式的数据集：`csv`, `tsv` & `json` 读入 `DataFrame`。

第二个步骤是评估，数据导入后通过输入命令查看数据集：`变量名`, `变量名.info()`, `变量名.head()`。这些操作可以帮助了解数据集的结构，同时通过对数据集的深入观察，在评估环节记录数据集中的脏数据和不整洁的数据。在这个步骤中只需描述发现的问题，暂时不需要去思考如何解决这些问题。

在第二步分析的基础上，对脏数据和不整洁数据进行清理。这个步骤一定要通过编程来清理而不是手动清理。因为不整洁数据是结构类的问题，先清理不整洁的数据，然后再清理脏数据。如果反过来，可能会产生无效的工作。清理过程中，每一个定义的问题在处理过后，都需要做一下测试，查看数据集已经按要求更改了。

第三步中将三个数据集合并成了一个。在清理完成后将这个新的数据集存为 `csv`。清理工作就完成了。

这个已经被清理完成的数据，可以通过 `Python matplotlib`, `R` 或 `Tableau` 发现数据集中隐藏的规律了。