

2011 年英国人口普查

2011年，英国对所有家庭进行了人口普查。我们将查看一个特定问题的数据：“KS608 职业”。

背景：英国地理

大不列颠及北爱尔兰联合王国 (UK) 是一个包含英格兰、威尔士、苏格兰（在撰写本文时）和北爱尔兰（但不包括爱尔兰共和国）的国家。该问题的人口普查数据来自国家统计局 (ONS)，仅适用于英格兰和威尔士。苏格兰和北爱尔兰有独立的统计局。

英格兰本身可以分为9个地区：

- 西北地区
- 约克郡和亨伯河
- 东北地区
- 西米德兰兹郡
- 东米德兰兹
- 西南
- 东部
- 东南部
- 伦敦

英国过去被进一步划分为县，但现在这些县的重要性要低得多。事实上，这里混合了县、单一当局和行政区——总共 152 个。这些都被称为“县级单位”（CLU）。

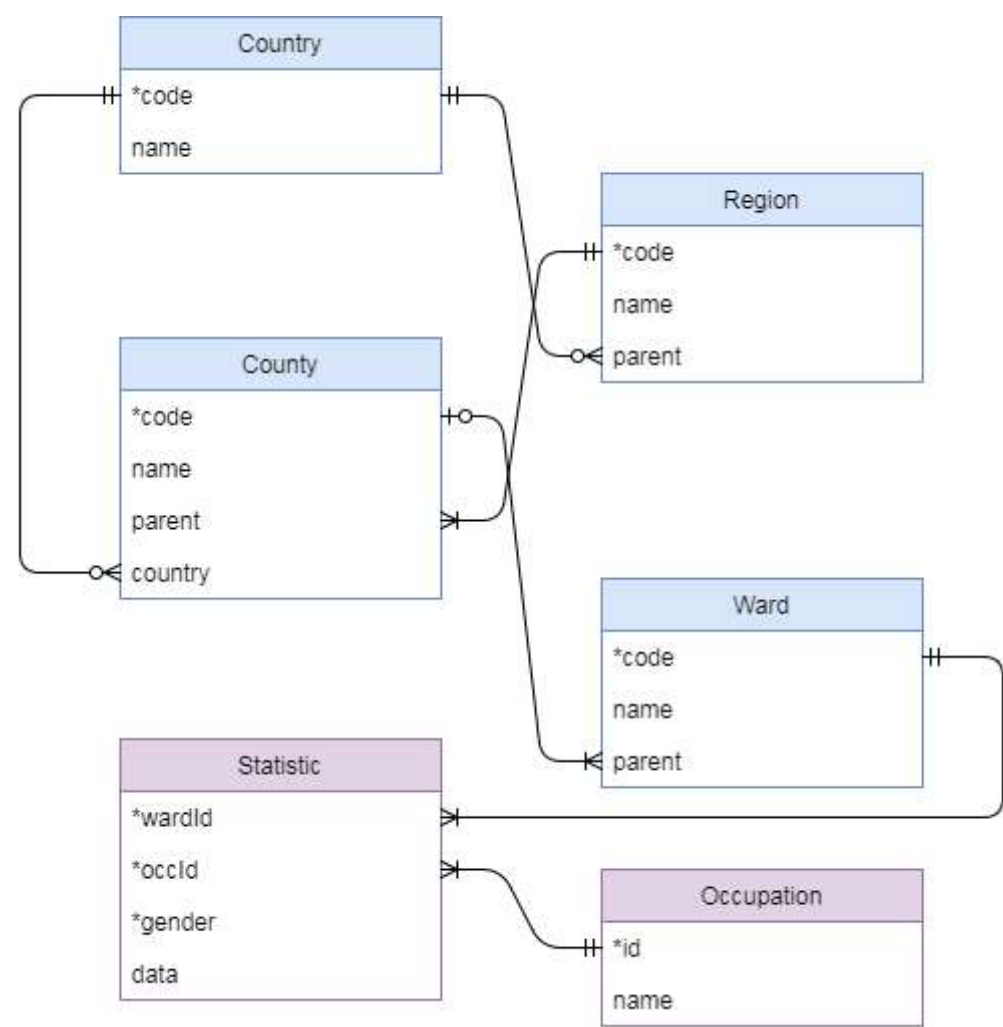
出于政治和统计目的，最小的相关单位是选举区，通常简称为选区。选区选举自己的议员（正如我们所见），并且可以在选区级别获得国家统计数据。2011 年人口普查时有 8570 个行政区。例如，布里斯托尔市统一管理机构包含35个选区，其中布里斯托尔大学位于卡伯特选区。

ONS 为每个统计单位（行政区、县、统一机构等）分配一个由 9 个字符组成的代码，Xaabbbbbbb 其中 X 代表英格兰，W 代表威尔士。前两位数字 aa 标识单位类型：05 为病房，06 为单一机关；在英格兰只有E12是地区，E08是自治市（大都会），E09是伦敦自治市，E10是县。最后 6 位数字标识各个单位。最后，所有英格兰和威尔士的代码为 E92000001 和 W92000004。

英国统计单位的交互式在线地图可在[UK data explorer](#)或[DataShine](#)在线获取。个别问题的所有人口普查数据都是公开的。（最有趣的数据——问题之间的相关性——由于隐私问题并不全部公开。研究人员可以向 ONS 申请访问特定数据集。）

数据格式

该 census 数据库具有以下架构。



对于英格兰的行政区， parent FK 指向县表并标识该行政区所属的 CLU（县级单位）。对于威尔士的选区，您在制定加入策略时需要考虑该 Ward.parent 列。 NULL

KS608：职业

2011 年人口普查的 KS608 问题要求所有年龄在 16 岁至 74 岁之间的就业英国公民将自己归类为 9 个职业类别之一。

- 查看数据库 Occupation 中的表格 census 可以看到 9 个职业类别。

问题的答案按 Statistic 以下格式记录在表中。性别为 1 表示女性，0 表示男性；在 2011 年人口普查中，这些是唯一的性别选择。

wardId	occId	gender	data
E05000001	1	1	54

该行记录在沃德 E05000001 (Aldersgate) , 54 名女性 (性别=1) 表示她们担任“经理、董事和高级官员” (occlId=1) 。

请注意 (在 ER 图中) 统计表的主键如何由三列 (wardId、occlId、性别) 组成。这正是您所期望的表, 其中每个行政区、职业类别和性别都有一个数据值。

练习

- 布里斯托大学位于该区 Cabot (区名并不总是明确的, 但这个是)。查找包含 Cabot 选区的 CLU、地区和国家/地区的名称和代码 (CLU = 县级单位 =“县表中的行”) 。
- 如果您在最后一个问题中使用了多个 SQL 查询, 请现在在一个查询中执行此操作。(换句话说, 找到适合您需要的表的联接策略。)
- 查找 Cabot 病房中职业类别 1 (经理等) 的女性人数。您可以直接使用在第一个查询中找到的 Cabot 病房代码和职业 id 1 - 此查询不需要任何 JOIN。
- 对于 Stoke Bishop 病房 (E05002003), 列出 9 个职业类别名称以及每个职业的男性人数。您的表应该有两列, 分别称为 name 和 number 。您可以使用提供的病房代码, 无需加入病房名称。

我们很快就能对人口普查数据进行更有趣的统计查询, 但为此我们需要 SQL 的统计函数, 我们将在下周学习这些函数。

最后还有一个稍微棘手的问题。仅使用我们迄今为止学到的技术就可以完成此任务。

- 找到所有不唯一的病房名称, 并按字母顺序打印它们 (每个仅一次) 。

总共有 400 个不同的此类名称 (例如, 有 21 个区称为“Abbey”) , 因此您的查询将生成一个相当长的表。您的查询可能还需要一段时间才能执行, 有更快的方法可以执行此操作, 但无法使用我们迄今为止学到的材料。该表以“Abbey、Alexandra、All Saints”开始, 以“Worsley、Wyke、Yarborough”结束。