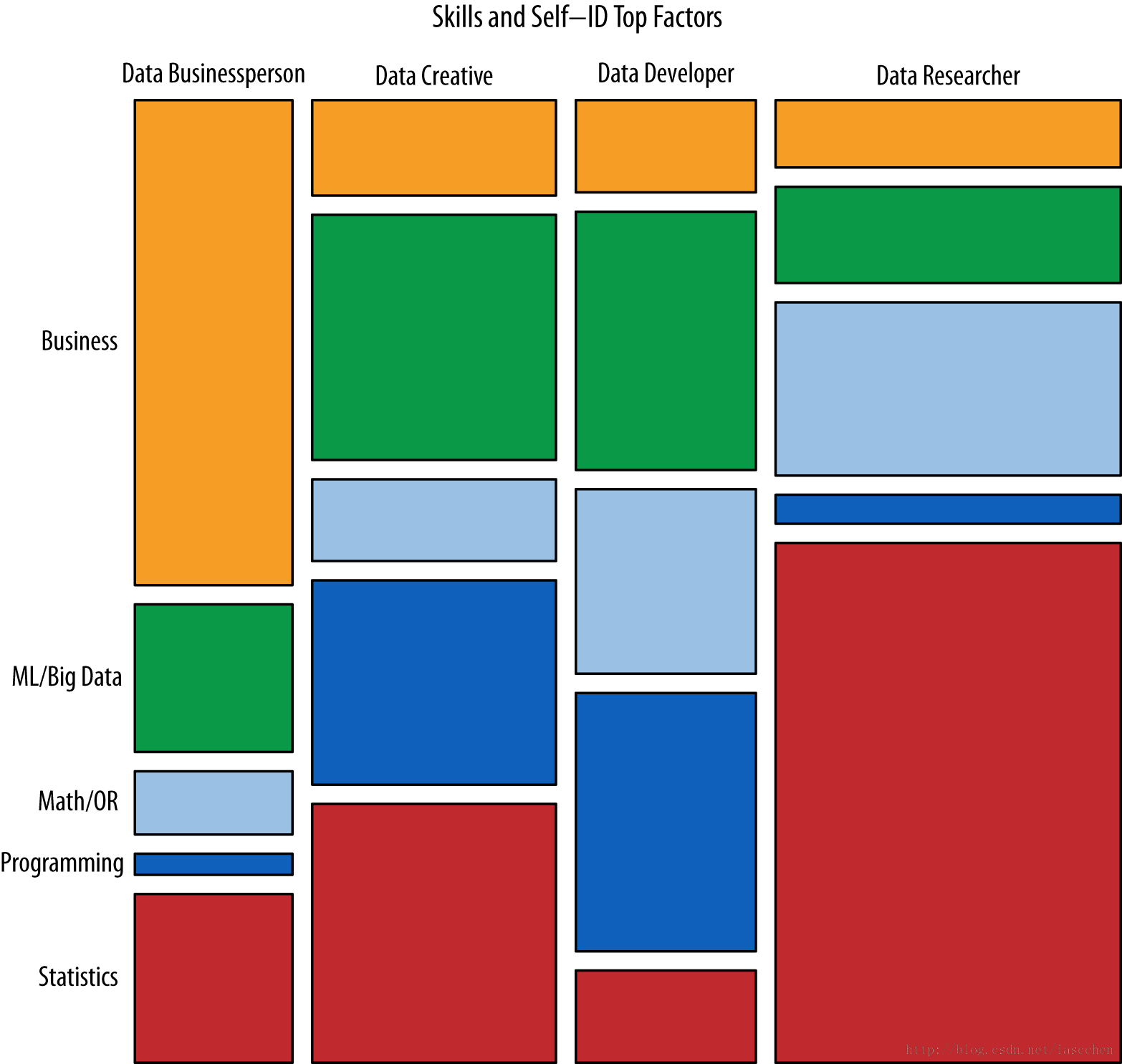
"Data Science = statistics who uses python and lives in San Francisco"

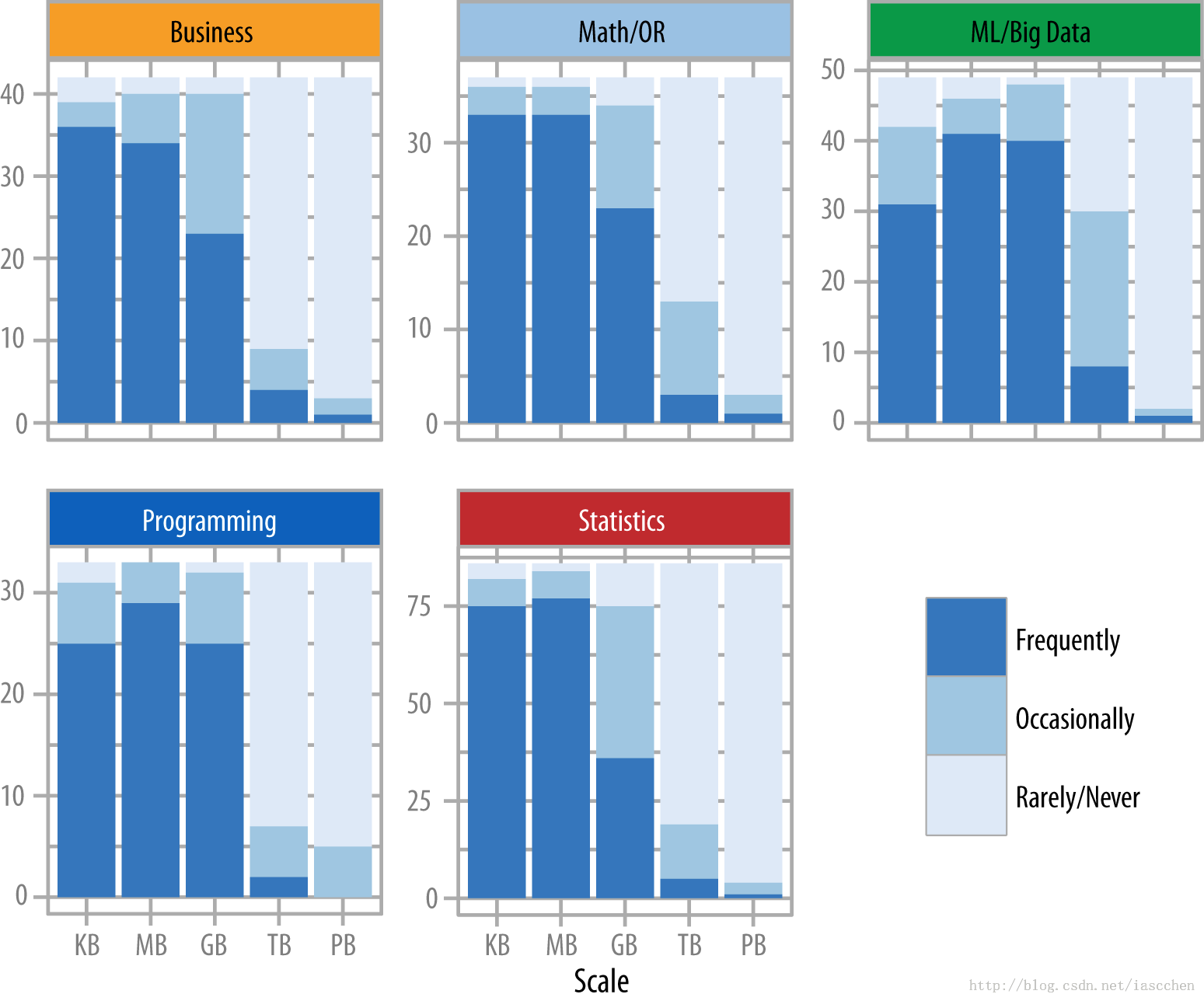
计算机科学家所谓的机器学习其实就是统计里面的prediction而已。

* report 告诉你过去发生什么
* BI告诉你现在发生什么
* modeling 告诉你将来发生什么





数据科学家分类

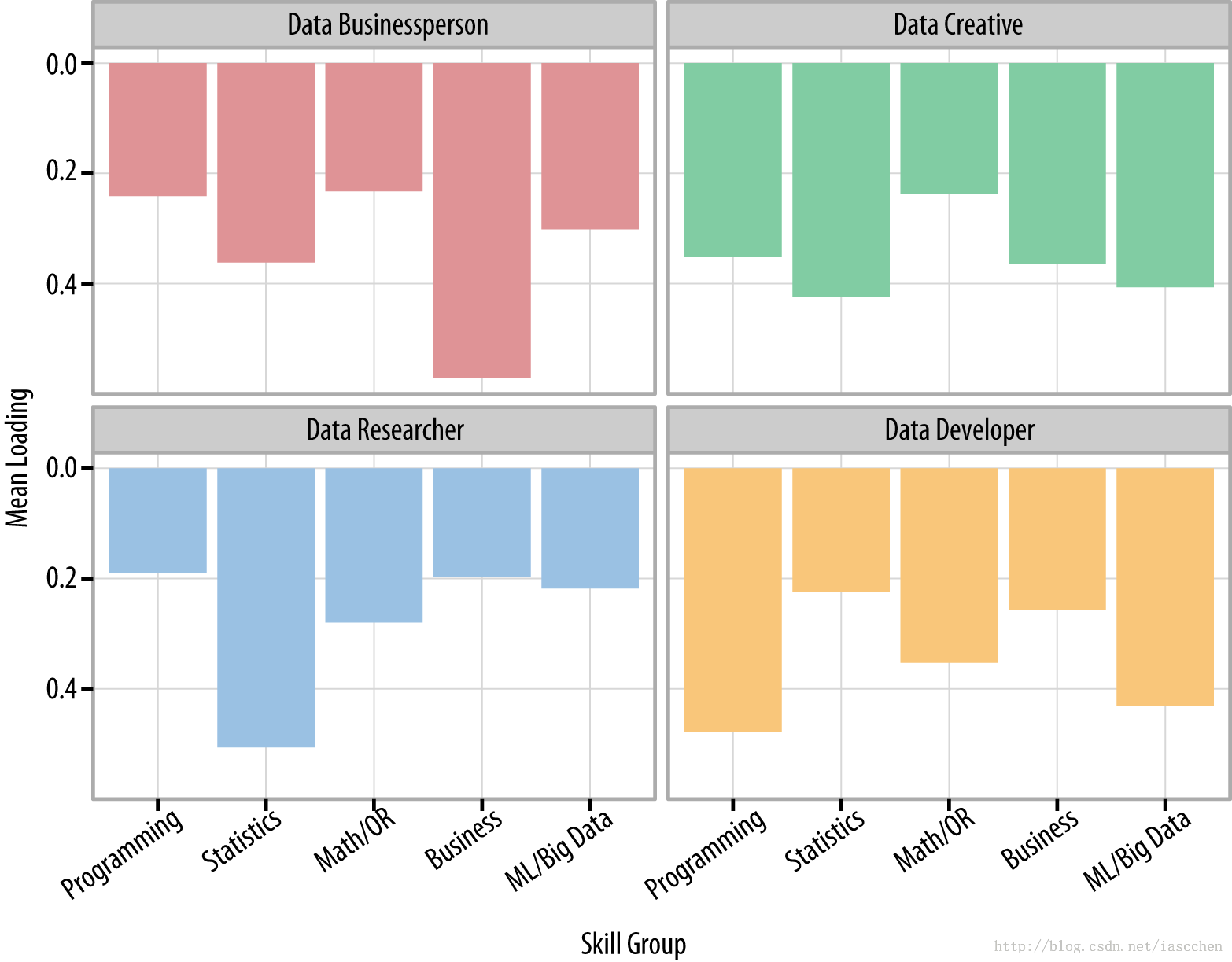


不同职位操作数据数量级大小

**T型数据科学家**

大多数成功的数据科学家既有丰富的专业经验，还至少在数据科学的某一方面深入钻研，如统计，大数据，或业务沟通等。这就是20世纪90年代提出的所 谓的“T形技能”模式：“T”字母顶部的横表示广博技能，而竖线表示在某个区域的深度钻研。

Data Businesspeople 具有顶级的商业能力，其他能力较弱。Data Researcher 往往具有非常深的统计相关技能，但一般说来，他们对于机器学习/大数据，商业，和编程技巧却相对低。Data Developer的模式可以被称为“π形”，拥有强大的编程技巧，较强的机器学习/大数据相关技能，而其他三个技能相对中庸。最后，Data Creative 是不像T形结构的，所有技能相对平均，既不最强，也不最弱。



“A data scientist is someone who is better at statistics than any software engineer and better at software engineering than any statistician.”

数据分析是一种靠谱的产品研究方法, 这玩意有很多误区, 也不能迷信, 最终到头来还是要人来做决策  
**忽略沉默的用户**

二战时英国空军为了降低飞机的损失，决定给飞机的机身进行装甲加固。由于当时条件所限，只能用装甲加固飞机上的少数部位。他们对执行完轰炸任务返航的飞机进行仔细的观察、分析、统计。发现大多数的弹孔，都集中在飞机的机翼上；只有少数弹孔位于驾驶舱。从数据上说, 加固机翼的性价比最高. 但实际情况缺恰恰相反, 驾驶舱才是最应加固的地方, 因为驾驶舱被击中的飞机几乎都没飞回来.

一个哥们，有次聚会时说自己买了很多条iPhone数据线，家里每个房间插一条，走到哪儿都能随时充电。。。。。在场的姑娘们都当成是段子笑笑就过去了， 只有个细心的姑娘悄悄问他买了多少条，他说42条。现在她和他在北京三环内总共42个房间的数套豪宅愉快地生活，下个月结婚。

【这就是数据分析】