

星战数据分析

数据来自于 [swapi.co](#)

基本统计

- 7部films
 1. The Phantom Menace*
 2. Attack of the Clones
 3. Revenge of the Sith
 4. New Hope
 5. The Empire Strikes Back
 6. Return of the Jedi
 7. The Force Awakens
- 87名characters
- 21颗planets
- 37艘starships
- 39架vehicles
- 37个species

字段汇总

Film: title、episode_id、opening_crawl、director、producer、release_date (1983-05-25) 、characters、planets、starship、vehicle、species、created (2014-12-18T10:39:33.255000Z) 、edited (2015-04-11T09:46:05.220365Z) 、url

Characters: name、height (cm) 、mass (kg) 、hair_color、skin_color、eye_color、birth_year (19BBY) 、gender、homeworld、films、species、vehicle、starships、created、edited、url

Planets: name、rotation_period、orbital_period、diameter、climate、gravity、terrain、surface_water、population、residents、films、created、edited、url

Starships: name、model、manufacturer、cost_in_credits、length、max_atmosphering_speed、crew、passengers、cargo_capacity、consumables、hyperdrive_rating、MGLT、starship_class、pilots、films、created、edited、url

Vehicles: name、model、manufacturer、cost_in_credits、length、max_atmosphering_speed、crew、passengers、cargo_capacity、consumables、vehicle_class、pilots、films、created、edited、url

Species: name、classification、designation、average_height、skin_colors、hair_colors、eye_colors、average_lifespan、homeworld、language、people、films、created、edited、url

对应关系

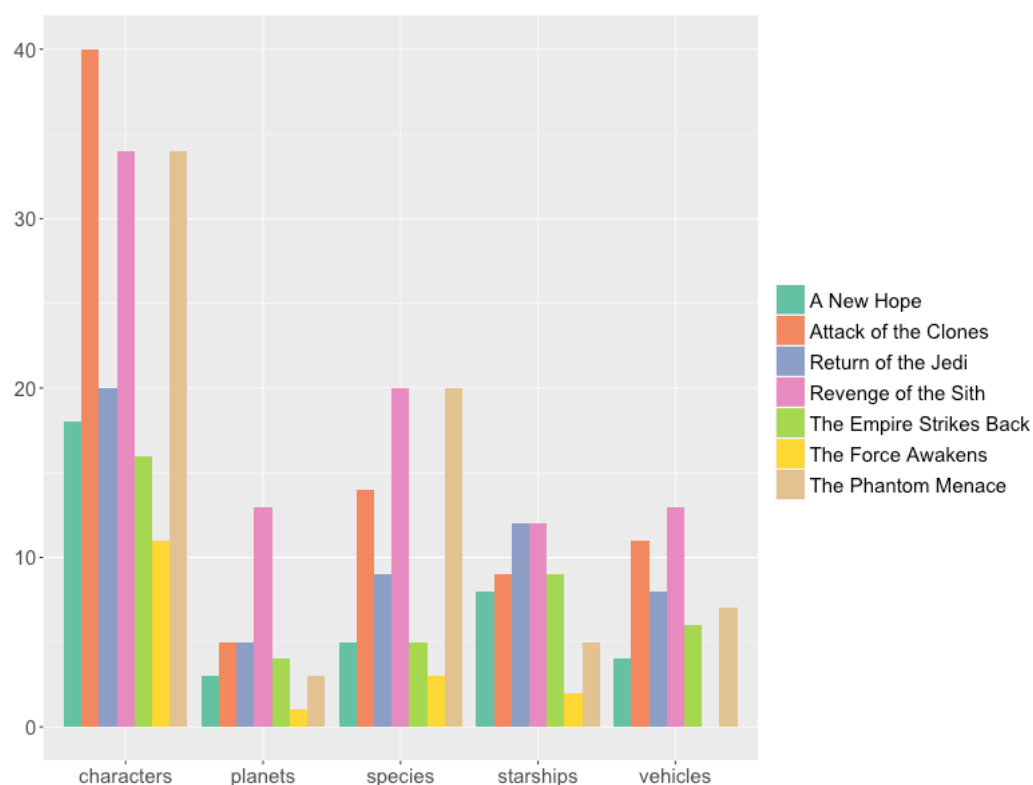
各类实体之间的关系如下:

	Films	Characters	Planets	Starships	Vehicles	Species
Films	-	M:M	M:M	M:M	M:M	M:M
Characters	M:M	-	M:1	M:M	M:M	M:1
Planets	M:M	1:M	-	-	-	1:1
Starships	M:M	M:M	-	-	-	-
Vehicles	M:M	M:M	-	-	-	-
Species	M:M	1:M	1:1	-	-	-

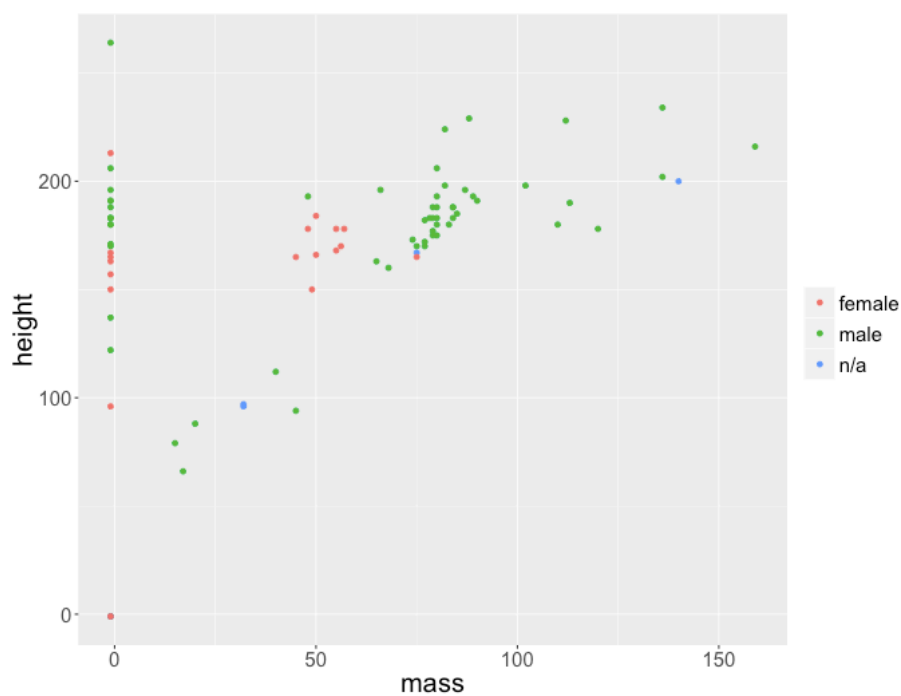
即film和其他实体都是多对多的关系， character和planet、species是多对一的关系， 而planet和species是一对一的关系。

基本统计

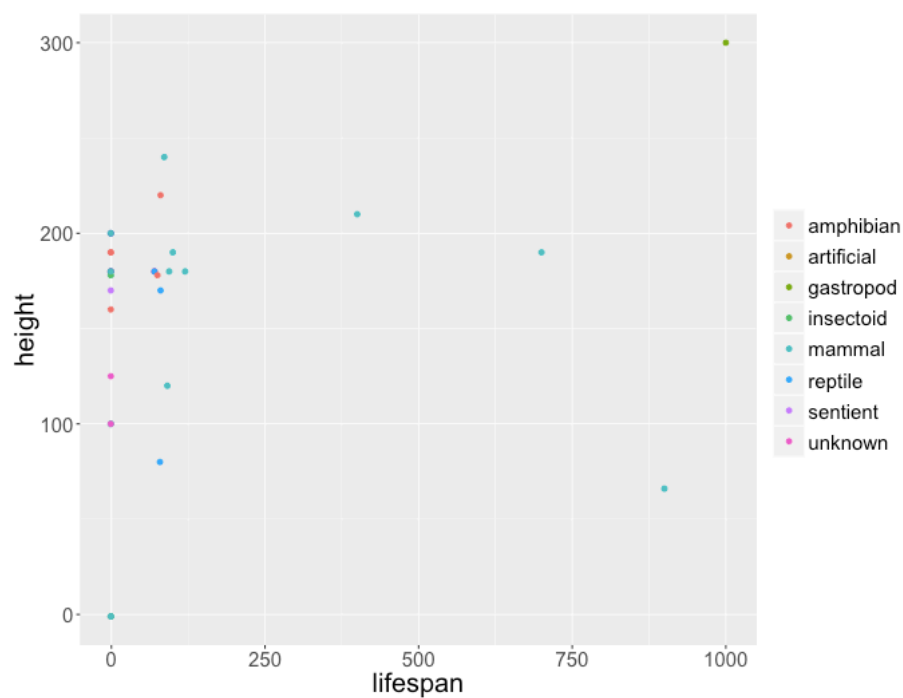
首先来看一下每部星战电影中涉及到了多少Characters、Planets、Starships、Vehicles和Species，如下图所示，可以看出*Attack of the Clones*参演人员最多， *The Force Awakens*涉及到的各类元素数量都比较少，当然也有可能是因为数据尚未完全整理的原因。



再来看一下Characters的一些统计信息，从下图中可以得出相关人物的体重和身高分布，体重或身高为-1表示值缺失，也有少数人物性别信息缺失。整体上身高和体重呈正相关，除了一个异常点（未在图中显示），Jabba Desilijic Tiure，175cm的身高，体重却达1358kg，而且为雌雄同体。



最后来看一下Species的一些统计信息，从下图中可以得出各个种族的平均身高和寿命，-1表示缺失值。大多数种族寿命在100年以内，有的可达几百甚至上千年，当然也有一个异常点未显示在图中，Droid，属于人造机器人，寿命为indefinite。



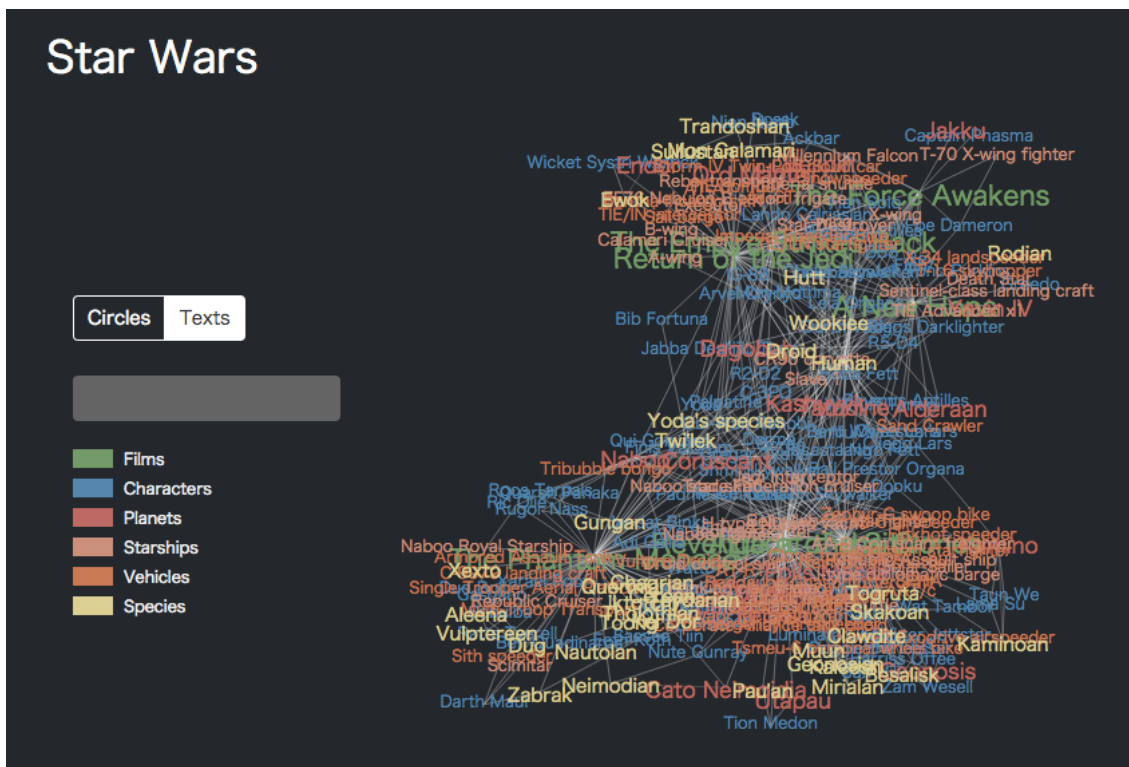
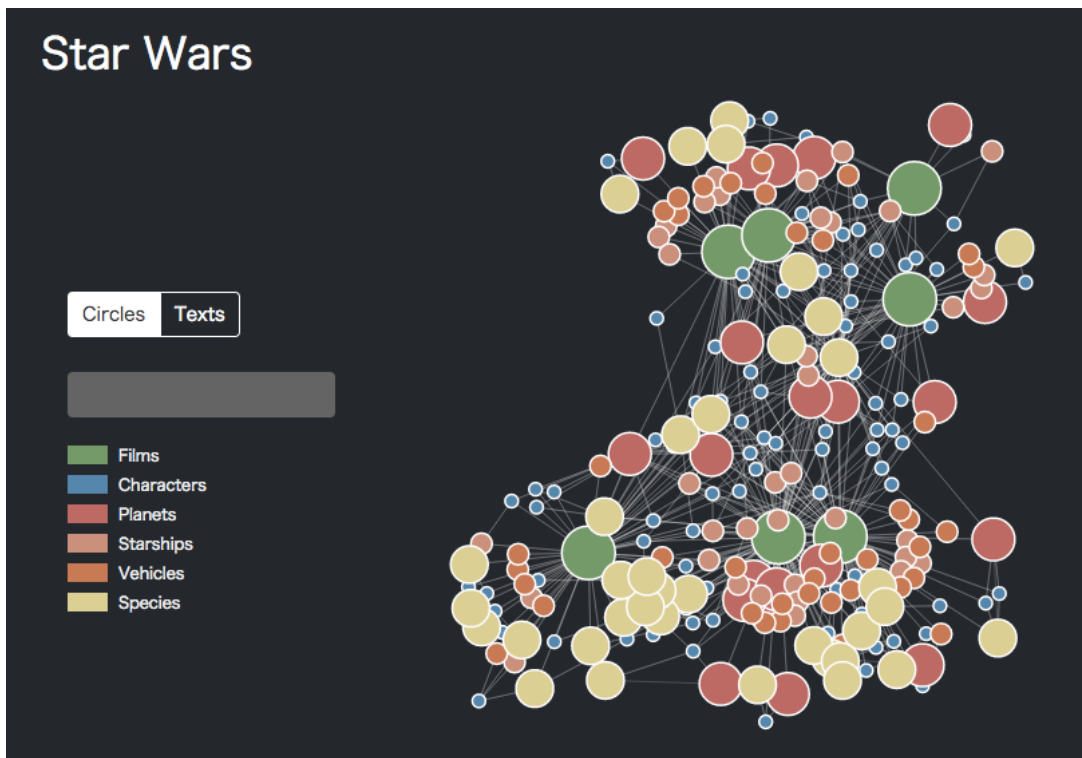
关系分析

七部星战涉及到了228个实体，这些实体之间的关联如何，可以使用知识图谱刻画它们之间的关系。在知识图谱中，节点表示实体、节点之间的边表示实体之间的关联，节点和边都可以包含自己的属性，因此知识图谱其实是个经丰富的图模型。为了生成一张知识图谱，需要完成以下三点内容：

1. 定义好有哪几类节点、节点之间存在哪几类关系、节点和边可能需考虑到哪些数据；
2. 基于准备好的数据，提取节点和边；
3. 使用可视化等技术进行展示，提供图谱探索和关联发现的功能。

在构建完知识图谱之后，我使用了D3进行展示。在这个知识图谱中，节点使用Circle或者Text来表示，使用不同的颜色来表示不同类别的节点，彼此相连的节点表示存在关联，如某个character参演了某部film、某个planet上生活着某个species等。鼠标悬浮在某个节点上时，仅显示和该节点之间相连的其他节点，从而能更清楚地看到节点之间的关联。

当然，也可以在搜索框中输入关键词，只有名称中包含关键词的节点才会显示，帮你快速找到感兴趣的节点。



时间线分析

知识图谱为知识表达提供了一种很好的解决方案，但是有些时候我们更关注实体的动态变化情况。举例来说，我们希望知道每部星战电影中，分别出现了哪些Characters、Planets、Starships、Vehicles和Species，显然这一答案无法从知识图谱中得出，所以接下来借助时间线进行可视化。

由于数据中并不包含时间戳信息，因此只能展示每个实体在哪些Films中出现过。如下图所示，不同的行代表films，不同的列代表不同的实体（entity），不同的颜色代表不同的实体类别。以演员为例，没有哪位演员参演过星战系列的全部电影。如果你对某个实体感兴趣，把鼠标悬停上去，就可以看到实体的名称啦！

