

Total: 221; Objects: 170; Place types: 51

Conclusions:

* Most often the intent word is a place type (e.g. What is the **region**) or an object (e.g., What is the **density**).
* Objects are used as intent three times as much as place types.
* Intent word is often part of an intent phrase (e.g., 2oo and 2ot).

# Using regular expressions to identify the intent word:

**Heuristic 1**: Any object or place type that occurs after auxiliary and immediately before the first relation (or verb, place name, date) in the question is an intent word.

**Example**: What is auxiliary the spatialadjective distributionadjective patternobject intent ofrelation

+ "<2::What[^>]\*>" # what question  
+ "<::(is|are|were|was|do|does|did|have|has|should|could|would|will)>"  
+ "(<::be>)?"  
+ "(<::the>|<::a>)?"  
+ "((<::and>)?(<::->)?<[otnqaspd]::(?P<adjective>[^>]+)>)\*" # zero or more adjectives  
+ "<[ot]::(?P<intent>[^>]+)>" # object or place type is an intent  
+ "("  
+ "<::" # a quick fix  
+ "|<[1-9rnqaspd]" # or followed by non-object and non-type code  
+ ")"

**Heuristic 2**: Any object or place type that occurs immediately before the auxiliary is an intent word.

**Example**: What residentialadjective buildingsplace type intent areauxiliary

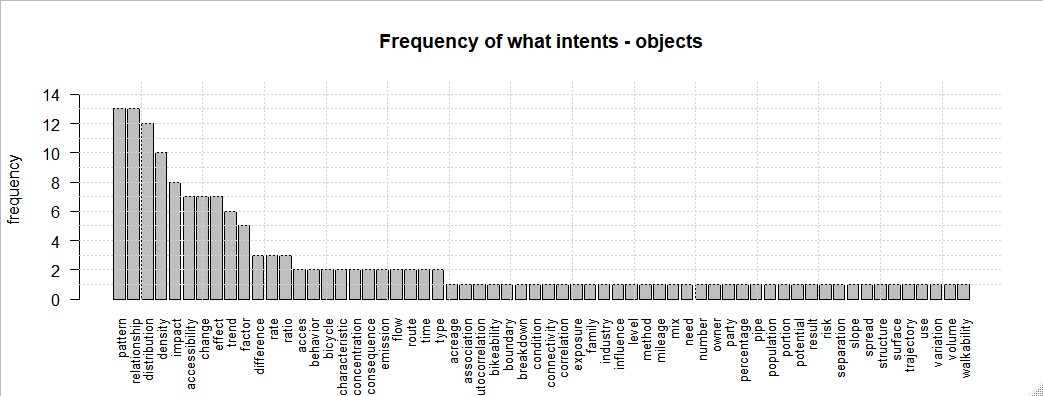
+ "<2::What[^>]\*>" # what question  
+ "((<::and>)?(<::->)?<[ntoqaspd]::(?P<adjective>[^>]+)>)\*" # zero or more adjectives  
+ "<[ot]::(?P<intent>[^>]+)>" # object or place type is an intent  
+ "<::(is|are|were|was|do|does|did|have|has|should|could|would|will)>" # either 'is' or 'are'  
+ "(<::be>)?"

|  |  |  |
| --- | --- | --- |
| Frequencies of matches to regular expression | | |
|  | Object intent | Place type intent |
| Heuristic 1 | 150 | 15 |
| Heuristic 2 | 7 | 32 |
| Total | 204 | |

Conclusion:

* Questions of the type “What is the [intent phrase]” are the most common.
* Questions of the type “What is the [intent phrase]” likely to have objects as intent word.
* Questions of the type “What [intent phrase] is” are more often about place types as intent.

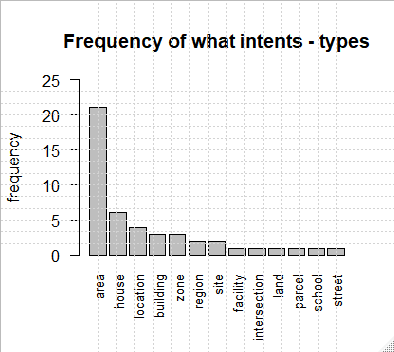
# Object intents:



Conclusion:

* Most often objects intents are about statistical measurements rather than thematic information. Maybe can be mapped to tools.
* It seems that the most frequent object intent can be mapped into the most common question categories identified from literature.
* 62% of all object intents is captured by 13 objects that occur 3 or more times.

# Place type intents:



Conclusion:

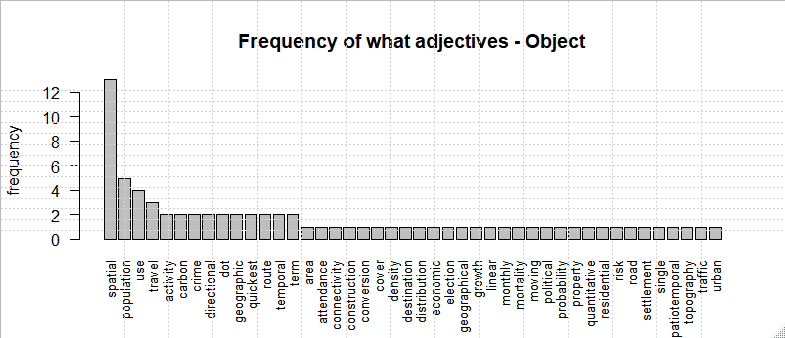
* Area is predominant place type intent in what questions.
* Most place types are either a geographical region or a building.
* 79% of all place type intents is captured by 5 place types that occur 3 or more times.

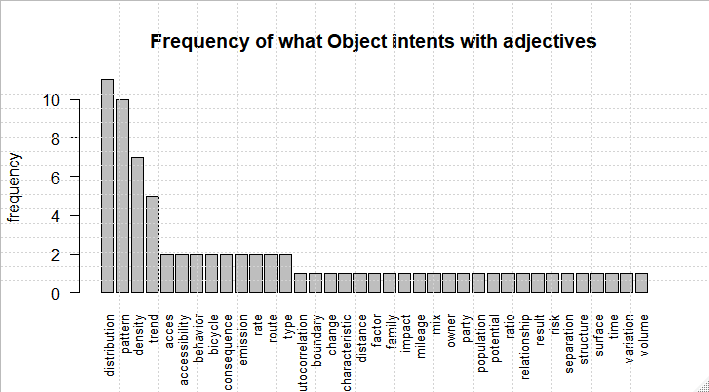
# Adjectives used with object intent:

**Examples**:

What residentialadjective buildingsplace type intent areauxiliary

What is auxiliary the spatialadjective distributionadjective patternobject intent ofrelation

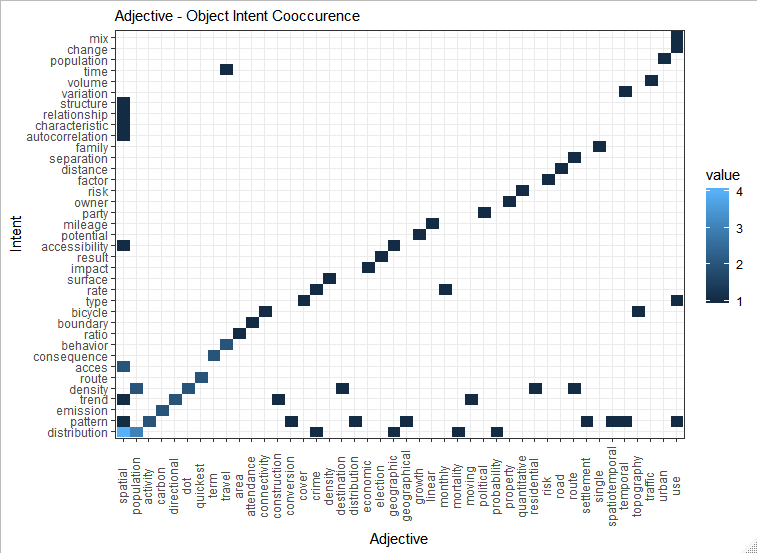




**Heuristic**: Only the adjective closest to the intent word is plotted in the graph below.

**Examples**:

What is auxiliary the spatialadjective distributionadjective patternobject intent ofrelation



Conclusion:

* The most frequent combinations of adjectives and object intents are **spatial distribution, population distribution**
* The most frequent adjectives are **spatial, population, land use, and travel**.
* Object intents that most frequently require adjectives are **distribution, pattern, density, and trend**.
* Adjectives can provide thematic context (e.g., election, crime, political).
* Intent words can also be used as adjectives for another intent (e.g., population, distribution).
* Why are some intent words require adjectives more frequently than others? Are they less self-descriptive, contain less information?n 5

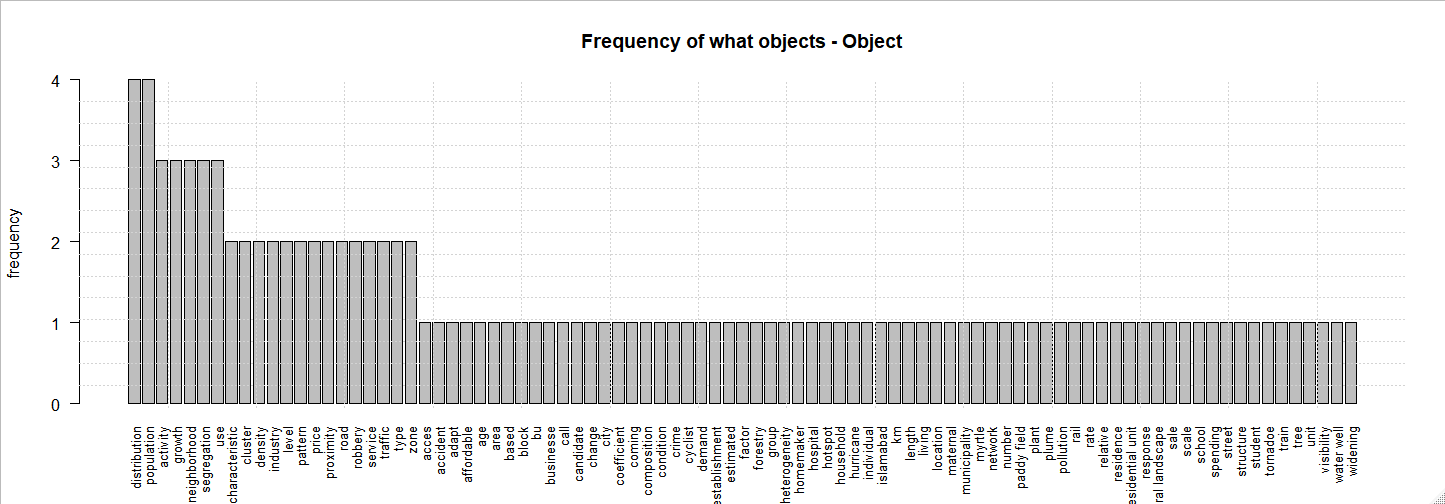
# Object of an intent:

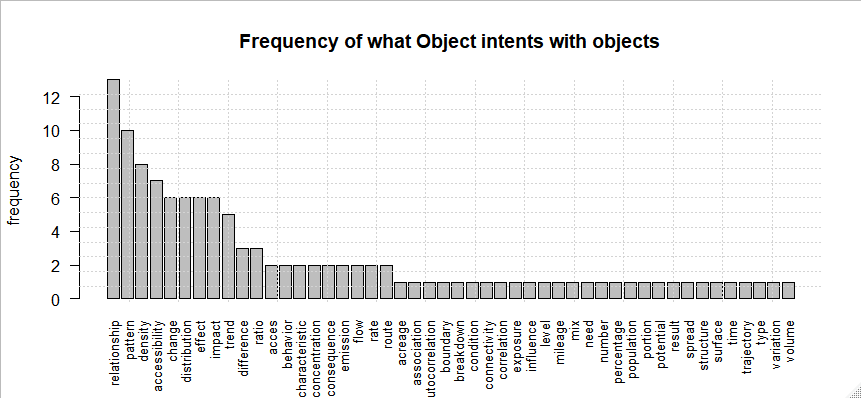
**Example**: What is the densityintent ofstart relation student populationobject of intent inend relation Utrecht.

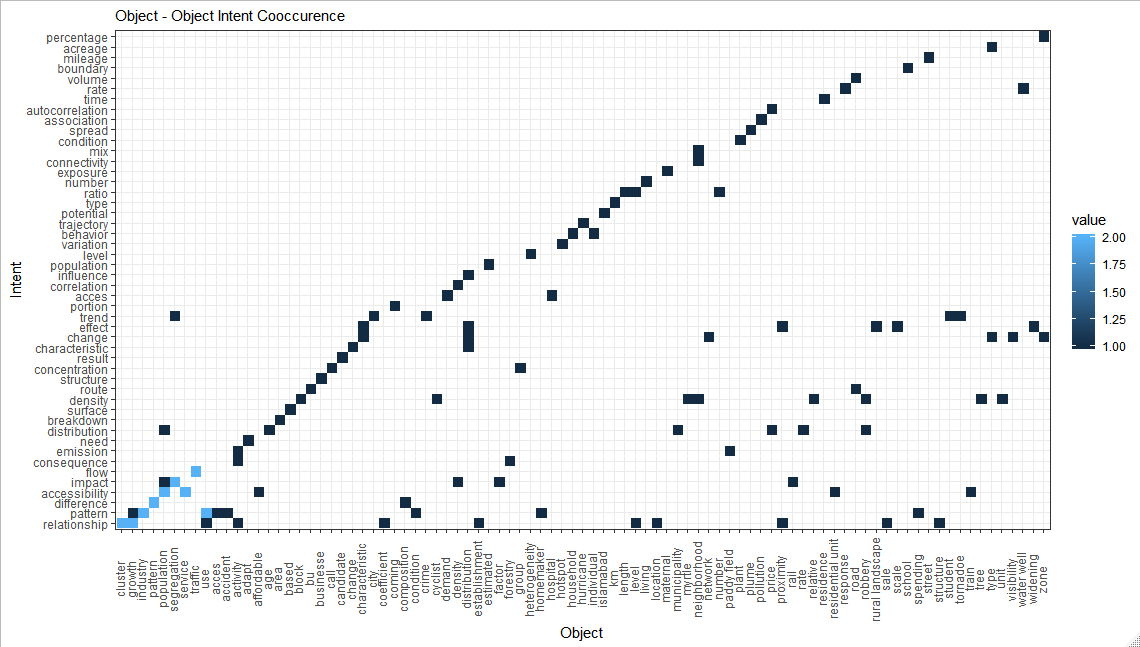
**Heuristic**: Object phrase of intent is always nested between two relations or between a relation and a clause. Object word is considered to be the last object in the object phrase.

+ "<2::What[^>]\*>" # what question  
+ "<::(is|are|were|was|do|does|did|have|has|should|could|would|will)>"  
+ "(<::be>)?"  
+ "(<::the>|<::a>)?" # 'the' is optional  
+ "((<::and>)?(<::->)?<[otnqaspd]::[^>]+>)\*"

+ f"<{intentCode}::(?P<intent>[^>]+)>"  
+ "(<r::|<::)(?P<relation>of|in|at|for|by|within|per|between|on|from|to|around|among|along|over)>"  
+ "(?!<(n|d)::)"  
+ "(?P<object>(.\*?))"  
+ "(?="  
+ "("  
+ "(<r::|<::)(of|in|at|for|by|within|per|between|on|from|to|around|if|among|along|that|over)>"  
+ "|<[1-9]::"  
+ ")"  
+ ")"





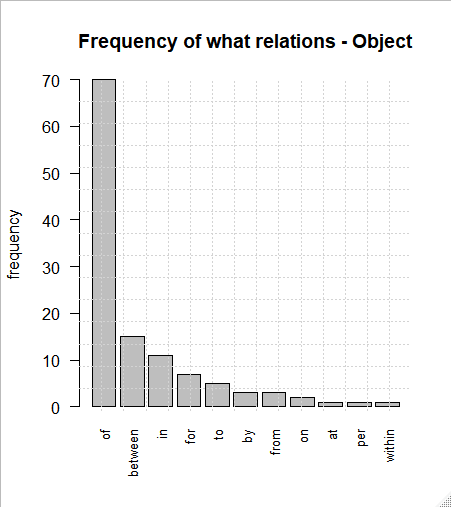


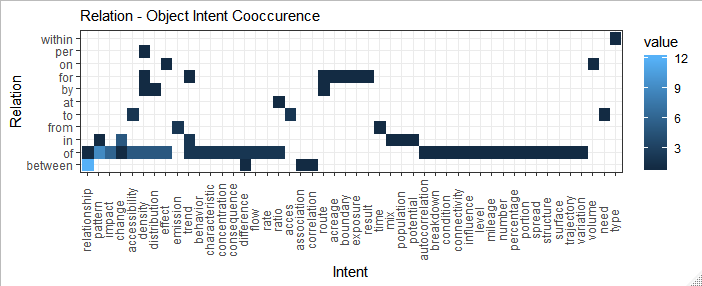
Conclusion:

* Same words used for intent can be used as object of intent (e.g., distribution, population).
* Similar to adjectives, objects of intent more often contain thematic information. Hence, there is the long tail in the distribution.
* Most frequent intents such as relationship, change, effect, impact involve interaction between two or more entities.
* Intents (such as pattern density, and distribution) that are likely to have adjectives are also likely to have objects? (but both at the same time or just adjective moved as object?)

# Relation connecting an intent and its object:

**Example**: What is the densityintent ofstart relation student populationobject of intent inend relation Utrecht.





Conclusion:

* Of is a universal relation that is used with most intents. So, it may not be very useful in extracting information.
* Between is a specific relation expressing relationship (relationship, difference, association, correlation) between two entities.