GI_Forum 2021, Issue 1
Page: 94 - 103
Research Paper
Corresponding Author:
@ tahira.ullah@uni-heidelberg.de
DOI: 10.1553/giscience2021_01_s94

Deriving Indicators for Points of Interest and Analyzing Mixed Activities in Urban Areas

Tahira Ullah[1], Sven Lautenbach[2] and Alexander Zipf[1,2]

[1]Heidelberg University, Germany
[2]Heidelberg Institute for Geoinformation Technology GmbH, Germany

Abstract

This paper evaluates indicators to analyse mixed activities, representing a combination of various facilities and services, in urban areas. Although mixed activities play an important role in urban planning projects, measuring themhas been problematic due to the lack of appropriate data and measurement approaches. In ecology, there are dozens of potential diversity indices, which have been deployed in recent land use studies to measure mixed activities. However, ecologists have highlighted that these indices are not always expressed inintuitive units. Recognizing the limitation of commonly used diversity indices, Hill numbers, which represent a mathematically unified family of diversity indices, are used. Taking advantage of new data sources such as Points of Interest (POIs) from OpenStreetMap, this study applied Hill numbers on POIs to measure mixed activities at a quarter level in Frankfurt. Results showed that Hill[1] (exponential of Shannon) is an appropriate quantitative measure to describe the diversity of facilities and services by a single numerical value. However, it is difficult to explain which factor, namely evenness or richness,has a stronger impact on the index. To gain a more comprehensive picture of mixed activities we suggest to considering further indicators such as evenness and richness.

Keywords: Diversity Indices, mixed activities, OpenStreetMap, Hill numbers, POIs

1 Introduction

The development of pedestrian-friendly city with a variety of services such as eating, education, healthcare, shopping ,and personal services has become an important goal for urban planning, local authorities, families, and economic groups (Manaugh & Kreider, 2013). There is an increasing demand from different parties to estimate the heterogeneity of facilities and services, also known as mixed activities, using an overall diversity indicator (Grant, 2002). Such an indicator could then be used to compare different city quarters, cities, or even different regions.

The measurement and assessment of biological diversity has a long history in ecology (Hill, 1973; Whitaker, 1965). Urban land-use studies have taken advantage of these biological indicators to analyse mixed activities. Two commonly used biodiversity indices that have been used in the context of land-use diversity is the Shannon index which quantifies the degree of mixture among different species (Frank, 1994; Manaugh & Kreider, 2013) and the Simpson's

Index (Ritsema van Eck & Koomen, 2008; Berger et al., 2004). Another promising indicator are Hill numbers,which integrate species richness, Shannon, and Simpson Index into a further class of diversity measures.

In this study, we explore the suitability of Hill numbers to assess the diversity of mixed activities based on Points of Interest (POIs) from OpenStreetMap (OSM). We will examine, how accurately mixed activities at a city quarter level can be measured, using the northern quarters of Frankfurt in Germany as a case study.

2 Study area: Data Source and Data Type

To measure mixed activities in urban areas we use Points of Interest (POI) from OpenStreetMap. OpenStreetMap (OSM) has the objective to create an open, free, digital map of the world through the efforts of volunteers (Goodchild, 2007). In OSM,POIs characterize important locations on a map represented by nodes, ways, or relations.

We obtained POI data for five city quarters in Frankfurt am Main via the Openpoiservice, which is a Web Service within the Openrouteservice infrastructure (Neis & Zipf, 2008). In total 55 categories of facilities were selected including eating, education, healthcare, shopping, and personal services. Indicators were calculated based on the POIs for each of the five city quarters separately.

For the study area, we selected the northern part of Frankfurt am Main due to its heterogeneity regarding mixed activities between the city quarters. Frankfurt am Main is the fifth biggest city in Germany with a population of 747.000 inhabitants (Stadt Frankfurt am Main, 2017). The quarters Eschersheim and Heddernheim comprise of various facilities and services due to their relative proximity to the city Center, whereas Praunheim and Niederursel are quieter quarters. Niederursel contains some of the scientific institutes of the Goethe-University Frankfurt. In Kalbach-Riedberg, one of the largest town-planning projects "Am Riedberg" was build, including residential areas, parks, green spaces, schools, healthcare services, and sports areas (Stadtplanungsamt Frankfurt am Main, 2021).

3 Methodology

Figure 1 provides an overview of the overall methodology. Facilities belong to different categories, which defines the level at which diversity indices are calculated. Since we had to assume that POIs from OSM are not complete and the proposed indices are sample size dependent, standardization was necessary to compare unequally large samples. We used the coverage estimator introduced by Good (1953). It is a measure of sample completeness, giving, in our context, the proportion of the total number of POIs in a quarter that belong to the categories represented in the sample (Chao & Jost, 2006). It describes the sum of the frequencies of the categories sampled, which is 100% when all categories are known (Chao, 2014). To estimate uncertainties, we used a bootstrap method, in which we calculated the Hill number and the coverage for each of the 100 bootstrap samples. Table 1 provides the average coverage of each city quarter and the respective reference sample size.

Table 1: Average coverage of quarters for reference sample n

City quarter	coverage	sample size *n*
Eschersheim	0.94	69
Heddernheim	0.96	125
Kalbach-Riedberg	0.94	70
Niederursel	0.98	89
Praunheim	0.95	51

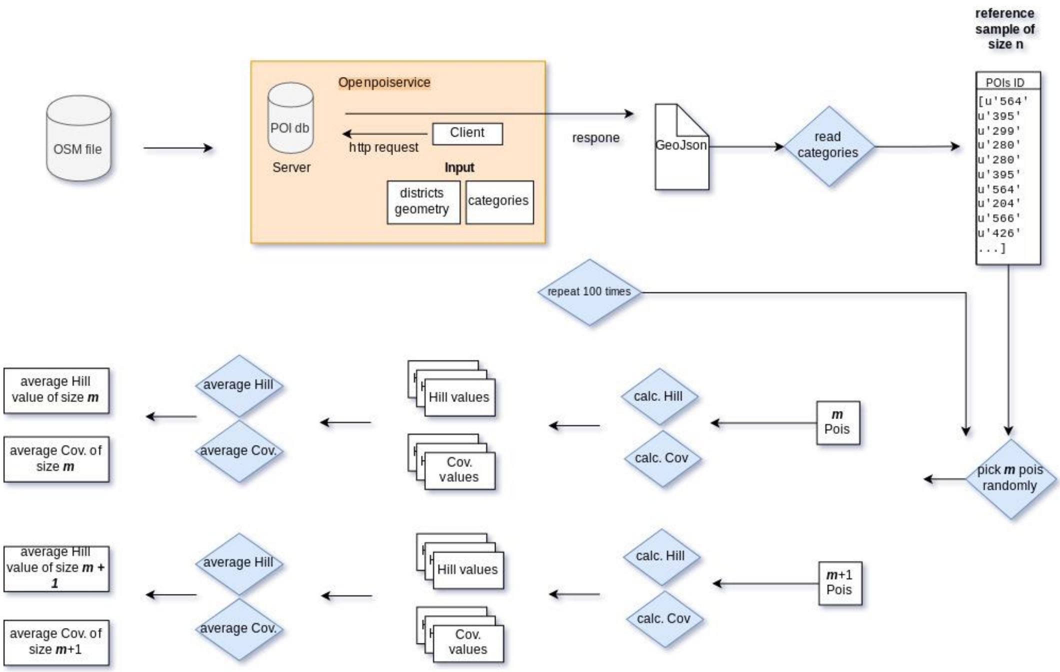

Figure 1: Workflow to calculate POI category diversity. For each city quarter Hill number and coverage were calculated. Uncertainty was estimated by a bootstrap approach.

While ecologists commonly refer to the terms species and individuals (Chao, 2014), in an urban context, we can consider categories as species and POIs as individuals. A city quarter is equal to an assemblage consisting of N total POIs, where each POI belongs to one of the C categories. Since we assume that the completeness of POIs in OSM vary from region to region, we consider a reference sample of n POIs from each city quarter from an underlying true assemblage that is unknown. The total number of categories observed in the reference sample is Cobs. Xi is the number of POIs of the ith category that is observed in the sample, i = 1, 2, …, C. Thus pi = Xi/n is the relative abundance of each observed category in the sample. Furthermore, fk is the number of categories represented by exactly k POIs, k = 0, 1, …, n. From the definitions above it follows:

$$n = \sum_{i=1}^{C} X_i = \sum_{k \geq 1} k f_k \quad (1b)$$

$$C_{obs} = \sum_{k \geq 1} f_k \quad (1a)$$

Species richness represents(in our case POI category richness)the simplest and yet most popular measure of diversity, describing the number of categories in a given area. However, one main limitation is that the observed categories are highly sensitive to the sample size. Furthermore, species or POI category richness does not take does not consider any information about the relative abundance of categories (Chao et al., 2014). Addressing the abundance problem, the Shannon and Simpson indices combine species richness and the relative abundance of each category into a single metrics. While the Shannon Index weighs each category exactly by their frequency, the Simpson Index is a dominance index that gives more weight to common or dominant categories (Ricotta, 2002; Jiang et al., 2017). However, the diversity indices are based on percentage composition, thus, they approach a constant value if sample size increases (Loya, 1972). Jost (2006) emphasises that the Shannon and Simpson are not necessarily themselves "diversity" indices; the Shannon index in particular represents an entropy reflecting the uncertainty in the outcome (Jost, 2006).

Hill numbers (Hill 1973) integrate species richness, Shannon, and Simpson index into a class of diversity measures. Thus, all measures include the following single expression for diversity:

$$D^q = \left(\sum_{i=1}^{C} p_i^q\right)^{1/(1-q)} \quad (2)$$

in which C is the number of categories in the assemblages, and the ith category has relative abundance p_i.The exponent q is also called the order of the diversity,thus,q determines the sensitivity to common or rare categories. The diversity of order 0 is completely insensitive to the relative frequencies of the categories,and is known as species richness (H^0) in ecology. If q is less than unity diversities favour rare categories, while all values of q greater than unity favour the most common categories (see Table 2). Order 1 is the exponential of Shannon entropy, while order 2 describes the inverse of the Simpson concentration (Chao et al, 2012).

Table 2: Conversion of common indices to Hill numbers

Index x:		Diversity in terms of x:	Hill numbers in terms of p_i:	Order:
Species richness	$x = \sum_{i=1}^{C} p_i^0$	x	$\sum_{i=1}^{C} p_i^0$	$Hill^0$
Shannon entropy	$x = -\sum_{i=1}^{C} p_i \cdot \ln(p_i)$	exp(x)	$\exp\left(-\sum_{i=1}^{C} p_i \cdot \ln(p_i)\right)$	$Hill^1$
Simpson concentration	$x = \sum_{i=1}^{C} p_i^2$	1/x	$1/\sum_{i=1}^{C} p_i^2$	$Hill^2$

Hill numbers offer the advantagethat they fulfil a doubling property. Therefore, if a city quarter is twice as diverse as another city quarter, the ratios of Hill numbers are always 2.00. Furthermore, traditional diversity indices can be converted to Hill numbers by simple algebraic transformation (see Table 2). Hill numbers are all expressed in units of effective number of categories, which are the number of equally abundant categories required to give the same value of diversity measure. Since Hill numbers have the same units, it is possible to graph them as a function of order *q* (see Fig. 2). The steepness of the curve reflects the evenness of a city quarter. Regarding Figure 2, if a city quarter has equally abundant categories,the curve is a constant at the level of $Hill^0$ (species richness). The ratios of Hill numbers can be used to obtain the evenness value as a single measure.

$$E_{a:b} = H^a / H^b \tag{3}$$

H_a and H_b are diversity numbers of order a and b based on *q*. Since $Hill^0$ (species richness) is highly dependent on sample size (see Fig. 3), we will use the ratio $Hill^2$ and $Hill^1$, which stabilizes with increasing sample size. The resulting value ranges between 0 and 1, if the quarter comprises of completely even distributed categories the evenness value is 1. In contrast, 0 means one category is dominating the whole quarter.

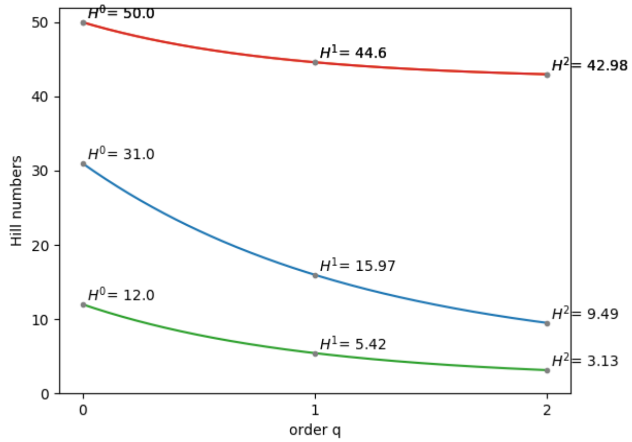

Figure 2: Hill numbers as a function of q. Steepness of the curve indicates, how even/uneven the distribution of the categories is. The red curve represents the most even city quarter, followed by the green curve. The blue curve indicates the most uneven distribution.

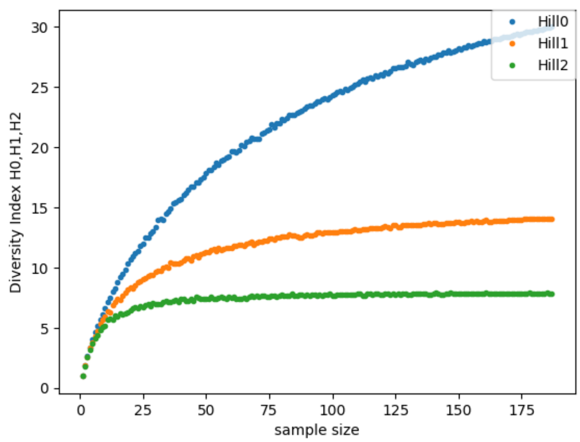

Figure 3: Hill numbers and sample size dependence. Hill0 (blue dots) is sensitive to sample size. Contrary, Hill1 (orange dots) and Hill2 (green dots) stabilize with increasing sample size.

As shown in Figure 3 Hill numbers, in particularly Hill0, is sample size dependent. Rarefaction describes an approach used in ecology to correct for this effect (Gotelli & Colwell, 2001). Rarefaction uses rarefaction curves to calculate category richness for a given number of individual samples. These curves plot the number of species as a function of the number of samples. They are based on a multiple resampling of the samples, and then plot the average number of species found in each sample (Gotelli & Colwell, 2001). However, if samples standardized by size are compared with each other, they will usually have a different degree of completeness. Chao & Jost (2012) suggest using a coverage-based standardization approach ensuring a comparison of samples of equal coverage. Based on these functional relationships the estimated Hill numbers are estimated.

4 Results

The estimated sample coverage values for the respective quarters were almost complete (see Table 1). To make the diversity of the five- quarter comparable, the respective quarters were rarefied down to the lowest base coverage of 94% (see dotted vertical line in Fig. 4). All three Hill numbers are evidently different for most city quarters. However, at 94% coverage the 95% confidence intervals of Hill^0 and Hill^2 for Eschersheim and Kalbach-Riedberg overlap, indicating that the POI diversity of these two city quarters is not significantly different.

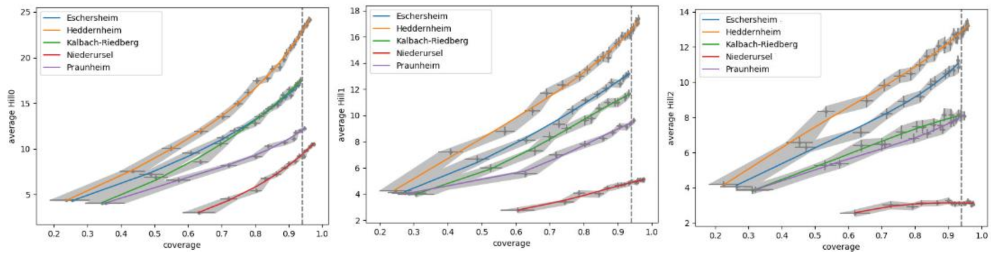

Figure 4: Sample completeness curves and Hill numbers (q=0, 1, 2) for the five city quarters. The grey areas indicate the 95% confidence bands derived from the bootstrap.

Table 3: diversity values(q=0,1,2) based on 0.94 basecoverage

City quarter	Hill^0	Hill^1	Hill^2	evenness	Sample size
Eschersheim	16.9	13.5	11.0	81%	69
Heddernheim	22.5	16.5	12.9	78%	125
Kalbach-Riedberg	17.3	11.8	8.0	68%	70
Niederursel	8.8	4.8	3.1	66%	89
Praunheim	11.4	9.6	8.0	83%	51

Jost (2006) points out that Hill^1 is a fair choice as a single diversity indexbecause it weighs categories exactly by their frequencies. Table 3 shows that Heddernheim has the highest diversity, followed by Eschersheim and Kalbach-Riedberg, while Niederursel has the lowest Hill^1 value. However, since Hill^1 focusses on both the richness and evenness, it is difficult to tell which factor contributed more simply by looking at the index. The richness (Hill^0) and evenness indicator provide additional information. Although Praunheim consists of few categories (11.4), the high evenness (83%) indicates an even distribution . Niederursel showed the lowest richness (8.8) and the lowest evenness (66%). Heddernheim consisted of the highest number of categories, hence showed the highest Hill^1 value. Eschersheim and Kalbach-Riedberg showed almost the same number of categories, but due to its lower evenness, Kalbach-Riedberg had a lower Hill^1 value.

5 Discussion

Biodiversity measures such as Species richness, Shannon's, and Simpson's index are quantitative measures reflecting the number of categories and their relative frequencies. While these widely used indices are non-linear and not sufficient to compare multiple areas, Hill numbers are more intuitive and meet a doubling property (Jost, 2006).

To compare the diversities of multiple quarters by a single numerical value, we suggest Hill1. However, Hill1 and Hill2 incorporates both richness and evenness, hence it is difficult to tell which factor contributed more simply by looking at the index. We recommend calculating a separate evenness measure in combination with species richness (Hill0), to derive further characteristics of the respective quarter.

Ecologists consider a site as diversewhen it consists of a high number of species,and if the species are well balanced in terms of abundance. When applying this terminology in an urban context, the following question arises: Are high richness and evenness sufficient for a diverse area in an urban context? Thus, the degree of diversity should always be considered with other factors, such as the population density and the resulting demand for specific services.

Moreover, diversity indices measure diversity quantitatively, they do not pay attention to qualitative aspects. For instance, a perfect mix of schools, grocery stores, and healthcare services scores identical to the same proportions of banks, estate agents, and companies.

In this study, we considered OSM POIs as a reference sample of an underlying true assemblage. However, if POI data is complete, standardization methods are unnecessarybecause the absolute number of POIs are also an indicator for the diversity of an area. Standardization methods are only useful when an area has not reached high completeness. Even though OSM-based POIs for urban areas in Germany seem almost complete, this work provided standardization methods so that the introduced methods are also applicable to areas with lower coverage. Overviews on methods for estimating completeness and other data quality indicators in OSM are given in . Degrossi et al. (2018), Ludwig et al. (2019), Barron et al. (2003) and related analysis frameworks like ohsome.org (Raifer et al 2019). For how those methods have been adapted for biodiversity see Jacobs & Zipf (2017).

6 Conclusion

In conclusion, this paper has discussed the application of biodiversity indices in an urban context. Evidently, a high diversity of services and facilities can support economic groups, families, urban planners, as well as local authorities to identify attractive areas or areas that lack a high diversity. The proposed diversity indicator can be used in regression analyses as a proxy to explain socioeconomic variations. Regarding the effectiveness of the introduced indices,further studies might focus on revealing associations between diversity and other relevant social factors.

References

Barron, C., Neis, P., & Zipf, A. (2014). A Comprehensive Framework for Intrinsic OpenStreetMap Quality Analysis. *Transactions in GIS*, *18*(6), 877–895. https://doi.org/10.1111/tgis.12073

Berger, A. N., Demirguc-Kunt, A., Levine, R., & Haubrich, J. G. (2004). Bank Concentration and Competition: An Evolution in the Making. *Journal of Money, Credit, and Banking*, *36*(3b), 433–451. https://doi.org/10.1353/mcb.2004.0040

Chao, A., Gotelli, N. J., Hsieh, T. C., Sander, E. L., Ma, K. H., Colwell, R. K., & Ellison, A. M. (2014). Rarefaction and extrapolation with Hill numbers: A framework for sampling and estimation in species diversity studies. *Ecological Monographs*, *84*(1), 45–67. https://doi.org/10.1890/13-0133.1

Chao, A., & Jost, L. (2012). Coverage-based rarefaction and extrapolation: Standardizing samples by completeness rather than size. *Ecology*, *93*(12), 2533–2547. https://doi.org/10.1890/11-1952.1

Degrossi, L. C., Albuquerque, J. P. de, Rocha, R. dos S., & Zipf, A. (2018). A taxonomy of quality assessment methods for volunteered and crowdsourced geographic information. *Transactions in GIS*, *22*(2), 542–560. https://doi.org/10.1111/tgis.12329

Frank, L. (1994). Impacts of Mixed Used and Density on Utilization of Three Modes of Travel: Single-Occupant Vehicle, Transit, Walking. *Transportation Research Record*, *1466*, 44–52.

Good, I. J. (1953). The Population Frequencies of Species and the Estimation of Population Parameters. *Biometrika*, *40*(3/4), 237–264. https://doi.org/10.2307/2333344

Goodchild, M. F. (2007). Citizens as sensors: The world of volunteered geography. *GeoJournal*, *69*(4), 211–221. https://doi.org/10.1007/s10708-007-9111-y

Gotelli, N. J., & Colwell, R. K. (2001). Quantifying biodiversity: Procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters*, *4*(4), 379–391. https://doi.org/10.1046/j.1461-0248.2001.00230.x

Grant, J. (2002). Mixed Use in Theory and Practice: Canadian Experience with Implementing a Planning Principle. *Journal of the American Planning Association*, *68*, 71–84. https://doi.org/10.1080/01944360208977192

Hill, M. (1973). Diversity and Evenness: A Unifying Notation and Its Consequences. *Ecology*, *54*, 427–432. https://doi.org/10.2307/1934352

Jacobs, C., & Zipf, A. (2017). Completeness of citizen science biodiversity data from a volunteered geographic information perspective. *Geo-Spatial Information Science*, *20*(1), 3–13. https://doi.org/10.1080/10095020.2017.1288424

Jiang, J., Shang, P., Zhang, Z., & Li, X. (2017). Permutation entropy analysis based on Gini–Simpson index for financial time series. *Physica A: Statistical Mechanics and Its Applications*, *486*(C), 273–283.

Jost, L. (2006). Entropy and diversity. *Oikos*, *113*(2), 363–375. https://doi.org/10.1111/j.2006.0030-1299.14714.x

Loya, Y. (1972). Community structure and species diversity of hermatypic corals at Elat, Red Sea. *Marine Biology*, *13*, 100– 123. https://doi.org/10.1007/BF00366561

Ludwig, C., Hecht, R., Lautenbach, S., Schorcht, M., & Zipf, A. (2019). *Assessing the Completeness of Urban Green Spaces in OpenStreetMap*. https://doi.org/10.5281/ZENODO.3387701

Manaugh, K., & Kreider, T. (2013). What is mixed use? Presenting an interaction method for measuring land use mix. *Journal of Transport and Land Use*, *6*, 63–72. https://doi.org/10.5198/jtlu.v6i1.291

Raifer, M., Troilo, R., Kowatsch, F., Auer, M., Loos, L., Marx, S., Przybill, K., Fendrich, S., Mocnik, F.-B., & Zipf, A. (2019). OSHDB: A framework for spatio-temporal analysis of OpenStreetMap history data. *Open Geospatial Data, Software and Standards*, *4*(1), 3. https://doi.org/10.1186/s40965-019-0061-3

Ricotta, C. (2002). Bridging the gap between ecological diversity indices and measures of biodiversity with Shannon's entropy: Comment to Izsák and Papp. *Ecological Modelling*, *152*, 1–3. https://doi.org/10.1016/S0304-3800(01)00468-9

Ritsema van Eck, J., & Koomen, E. (2008). Characterising urban concentration and land-use diversity in simulations of future land use. *The Annals of Regional Science*, *42*(1), 123–140. https://doi.org/10.1007/s00168-007-0141-7

Stadt Frankfurt am Main (2017). Statistisches Jahrbuch Frankfurt am Main. Retrieved from https://frankfurt.de/service-und-rathaus/zahlen-daten-fakten/publikationen/statistisches-jahrbuch

Stadtplanungsamt Frankfurt am Main (2021). Am Riedberg. Projektbeschreibung. Retrieved from https://www.stadtplanungsamt-frankfurt.de/am_riedberg_5309.html

Whittaker, R. H. (1965). Dominance and Diversity in Land Plant Communities: Numerical relations of species express the importance of competition in community function and evolution. *Science (New York, N.Y.)*, *147*(3655), 250–260. https://doi.org/10.1126/science.147.3655.250