Analisis Data Biologi dengan RStudio

Hardskill Share Workshop Himabio NYMPHAEA

Sabhrina Gita Aninta

Queen Mary University of London

2021/10/31 (updated: 2021-11-12)

Yang akan kita lakukan hari ini

Diskusi prinsip dan alur analisis data menggunakan R[1]

Latihan analisis data keanekaragaman burung dan vegetasi[2]


ECOGRAPHY

Research

Diverse temperate forest bird assemblages demonstrate closer correspondence to plant species composition than vegetation structure

Bryce T. Adams and Stephen N. Matthews

B. T. Adams (http://orcid.org/0000-0002-7830-6521) (adams.861@osu.edu) and S. N. Matthews, School of Environment and Natural Resources, The Ohio State Univ., Columbus, OH 43210, USA.

Ecography
42: 1752–1764, 2019
doi: 10.1111/ecog.04487
Subject Editor: Cagan Sekercioglu
Editor-in-Chief: Miguel Araújo
Accepted 22 June 2019

The aggregate structure of vegetation has long provided a strong conceptual basis for understanding the ecological separation of faunal species, while correspondence with plant species composition remains largely underdeveloped and considered secondary to structure. Longstanding ecological debate on the matter has likely been sustained by statistical methods incapable of accommodating an entire species composition in the explanatory role. We used direct ordination methodologies (predictive co-correspondence and canonical correspondence analyses) that allow the comparison of prediction levels between composition and structure for understanding avian assemblage composition in a temperate forestland in southeastern Ohio. Compositional (birds and woody plants) and structural (both vertical and horizontal dimensions) data were collected from point samples comprising a spectrum in topography and successional state. Total woody plant composition (11.22%) explained more (cross-validatory) variation in avian species composition than structure, quantified with dense LiDAR recordings (7.35%) and field methods (6.31%). Plant composition assessed an integrative character, synthesizing aspects of environmental condition, structure, and most likely species-specific preferences for tree and shrub species that structural indices alone could not for predicting avian species composition. These results conflict with the traditional view that structure be most influential to avian assemblage composition. Instead these results demonstrate that plant and avian assemblages are closely linked, and that plant species per se can be a powerful tool for predicting avian habitat. More importantly, in furthering ecological understanding, it is critical to consider the complex web of interacting processes that make up temperate forest ecosystems, in which composition occupies a key position.

Keywords: birds, co-correspondence analysis, composition versus structure, MacArthur, temperate forests

NSO NORDIC SOCIETY OIKOS
www.ecography.org

[1] BES. 2017. Guide to Better Science: Reproducible Code. British Ecological Society. https://www.britishecologicalsociety.org/publications/guides-to

[2] Adams and Stevens. 2019. “Diverse temperate forest bird assemblages demonstrate closer correspondence to plant species composition than vegetation structure.” Ecography. https://doi.org/10.1111/ecog.04487

Prinsip alur analisis yang baik

- Mulai dari salinan data mentah (**data mentah tidak boleh berubah**)
- Seluruh pemrosesan data, baik *cleaning*, *filtering*, dsb harus melalui skrip, tidak boleh manual di dalam file data
- Pisah alur skrip R ke beberapa unit tematis.
    - Contoh: bagian (1) membaca dan membersihkan data, (2) analisis data, (3) produksi tabel dan grafik
- Jika kamu melakukan beberapa kode lebih dari satu kali, buatlah `function`, dan dokumentasikan dengan baik apa yang dilakukan oleh kode tersebut: apa input dan outputnya, apa yang dilakukan fungsinya, dan mengapa
- Dokumentasikan kode dan data melalui komentar
    - menggunakan '#' dalam skrip, atau
    - dokumentasi tertulis yang terpisah
- Segala luaran perantara dalam alur analisis **harus dipisah dari data mentah**

Yang biasa saya lakukan

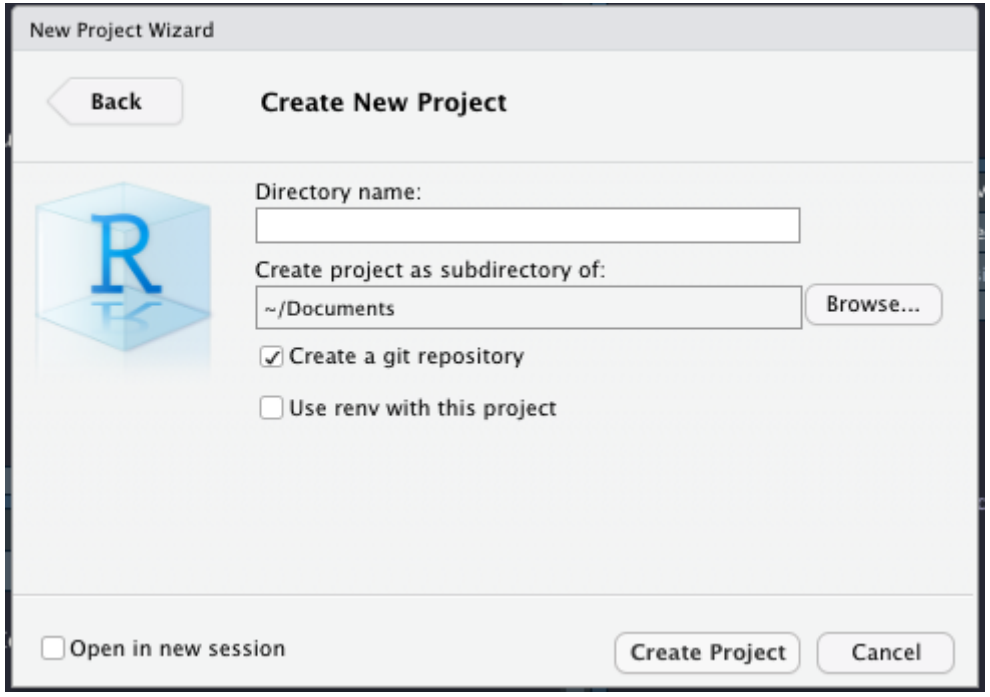
1. Menginisiasi `Project` dalam RStudio
2. Membuat struktur *folder* yang `Project`
3. Tulis `README` dalam folder `Project`
4. Membaca data dan mengecek bentuk data
5. Mengubah bentuk data sesuai tujuan analisis
6. Analisis data dengan eksplorasi (statistik deskriptif)
7. Uji statistik

"Embrace a one folder = one project mentality. Rstudio's "R projects" are excellent for encouraging this. This habit enables easy communication with other scientists and for that reason it is so important."

Anonymous

Membuat R Project

Untuk membuat R Project, klik `File` -> `New Project...` -> `New Directory`



- Tulis nama folder `Project`
- Pilih `directory` yang diinginkan kalau belum sesuai
- Klik `Create Project`.

Nanti akan muncul tampilan R Studio kosongan.

Membuat struktur folder dalam project

1. Konsisten
    - Apa pun sistem penamaan yang digunakan, pastikan untuk tetap menggunakan sistem tersebut
2. Hierarkis
    - Mulai dari jenis folder paling dasar, lalu buat folder anakan sesuai keperluan.
3. Gunakan README untuk mendeskripsikan struktur folder
4. Segala data dari koleksi, entri, atau mungkin kolaborator perlu "dikarantina" dan tidak disentuh
5. Catat semua alur ide, diskusi, dan keputusan analisis.
    - Buat folder `info` jika perlu untuk hal seperti ini
6. Pisahkan pekerjaan yang sudah selesai dan sedang berlangsung.

Contoh struktur folder untuk analisis sederhana

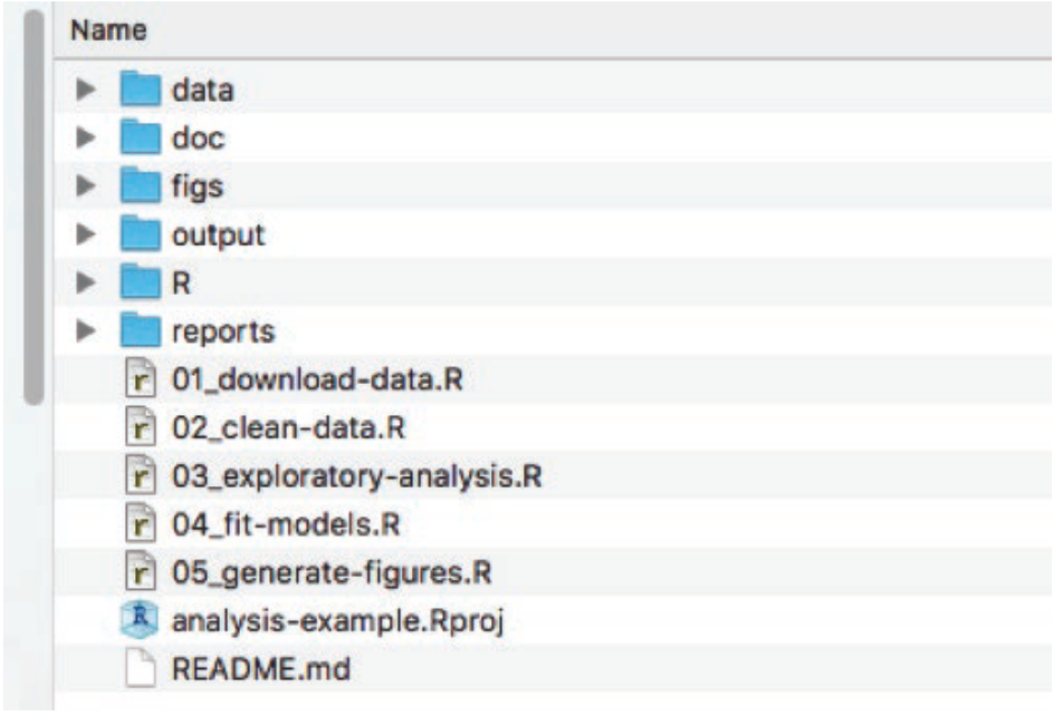


Fig 1. Example file structure of a simple analysis project

[*] diambil dari Guide to Better Science: Reproducible Code. British Ecological Society

"*Never ever touch raw data. Store them permanently, and use scripts to produce derived, clean datasets for analyses.*"

Francisco Rodríguez-Sánchez, Estación Biológica de Doñana (CSIC)

Beberapa cara menamai file di R

- *Machine readable*
  - Hindari spasi, tanda baca, huruf beraksen.
  - Gunakan penisah seperti "_" untuk memisah info file
- *Human readable*
  - Pastikan nama file juga mengandung info tentang file
- **Mudah diurutkan**
  - Bisa dengan memulai dengan angka atau tanggal.
  - Kalau pakai tanggal, gunakan format YYYY-MM-DD
  - Untuk skrip, bisa menomori urutan penggunaan skrip. Contoh: 1_data-cleaning.R, 2_linear-model.R

Contoh penamaan file yang baik dan buruk

Examples of bad vs. good filenames

BAD	BETTER
`01.R`	`01_download-data.R`
`abc.R`	`02_clean-data_functions.R`
`fig1.png`	`fig1_scatterplot-bodymass-v-brainmass.png`
`IUCN's metadata.txt`	`2016-12-01_IUCN-reptile_shapefile_metadata.txt`

[*] diambil dari Guide to Better Science: Reproducible Code. British Ecological Society

Struktur folder dan sistem penamaan yang bagus membuat alur analisis mudah dirunut

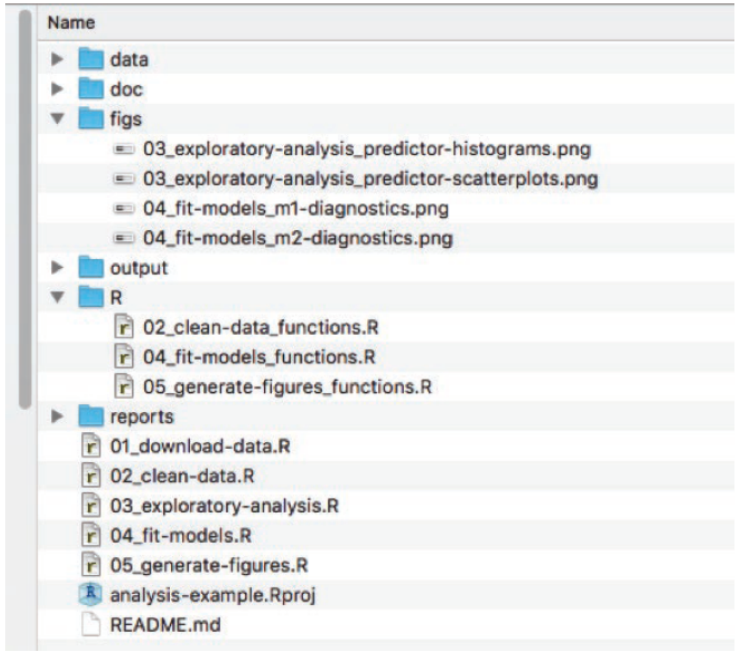


Fig 2. Linking analysis associated files (e.g. R scripts with functions) and outputs (generated figures) through the use of consistent naming.

[*] diambil dari Guide to Better Science: Reproducible Code. British Ecological Society

Prinsip analisis dengan bahasa pemrograman

1. Gaya penulisan kode logis dan mudah dibaca

2. Banyak komentar dan dokumentasi di setiap tahap

3. Tulis kode yang jika dipindah ke komputer lain bisa langsung dijalankan tanpa masalah (*portable code*)

4. Buat `function` untuk beberapa tahap yang terus menerus diulangi

Bisa dibaca lebih banyak lagi di *Guide to Better Science: Reproducible Code*. British Ecological Society (lihat link di slide pertama)

Demo: Apakah keanekaragaman burung lebih dipengaruhi struktur atau komposisi spesies vegetasi?

Analisis data Adams & Matthews (2019)

Tentang Data

Dalam latihan ini kita akan membaca data komunitas burung dan vegetasi dari bentang alam hutan campuran di Ohio, Amerika Serikat.

Spesies burung dan vegetasi diperoleh dengan metode *point count* di 210 titik sebanyak 6 kali survei sepanjang tahun 2015-2016.

Data tersebut diambil dari tiga tipe habitat:

1. *Ridgetops and southwestern hillslopes* atau *dry-oak* (do) atau habitat kering, 96 titik
2. *Northeastern hillslopes* atau *dry-mesic* (dm) atau habitat campuran, 59 titik
3. *Bottomlands* atau *wet-mesic* (wm) atau habitat riparian, 55 titik

Membaca data dan mengecek bentuk data

```
# Unggah library yang diperlukan --------
library(vegan)
library(tidyverse)

# Baca data -------
bird<-read.csv("input/bta_snm_vbirds.csv")
veg<-read.csv("input/bta_snm_vallde.csv")
veg_str<-read.csv("input/bta_snm_vfield.csv")
```

Perhatikan bahwa di sini dicontohkan pula:

- cara memisah sub-bagian skrip dengan karakter '-'
- cara mengacu direktori anakan dalam suatu R `Project`
- cara mengomentari tahapan kode

Membaca data dan mengecek bentuk data

```
# Tunjukkan tiga baris pertama data
head(bird,3)
```

```
##   X ACFL AMCR AMGO AMRE AMRO BAOR BAWW BGGN BHCO BLBW BLJA BRTH BTNW BWWA CACH
## 1 1    0    0    1    0    0    0    1    0    0    0    1    0    0    0    0
## 2 2    9    0    0    0    3    0    2    2    1    0    1    0    0    0    0
## 3 3    4    0    0    0    3    0    0    4    4    0    0    0    0    0    0
##   CARW CEDW CERW CHSP COYE EAPH EATO EAWP ETTI FISP GCFL GRCA HOWA INBU KEWA
## 1    0    0    0    0    0    0    1    2    0    0    0    0    7    0    0
## 2    0    0    0    0    0    1    0    0    0    0    0    0    4    0    0
## 3    0    0    3    0    0    0    1    0    2    0    0    0    2    0    0
##   LOWA NOCA NOPA OVEN PIWA PRAW RBGR REVI RWBL SCTA SUTA WBNU WEVI WEWA WOTH
## 1    0    0    0    6    0    0    0   12    0    0    0    1    0    0    0
## 2    0    1    0    6    0    0    0    6    0    4    0    0    0    2    7
## 3    0    0    0    8    0    0    0    9    0    3    0    3    0    0    6
##   YBCH YTVI YTWA
## 1    0    1    0
## 2    0    1    0
## 3    0    4    0
```

Membaca data dan mengecek bentuk data

```
# Melihat dimensi data (baris dan kolom)
dim(veg)
```

```
## [1] 210  66
```

```
dim(bird)
```

```
## [1] 210  49
```

```
# Tunjukkan tiga baris pertama data dan 5 kolom pertama
veg[1:3,1:5]
```

```
##   X Acer_negundo Acer_rubrum Acer_saccharinum Acer_saccharum
## 1 1            0   127.32395                0       633.9375
## 2 2            0   265.77093                0       127.3240
## 3 3            0    87.24917                0       290.6993
```

Membaca data dan mengecek bentuk data

```
# Tunjukkan tipe tiap variabel yang ada dalam data
str(veg_str)
```

```
## 'data.frame':    210 obs. of  6 variables:
##  $ X           : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ all_stem_den: num  11017 3200 7975 15294 17181 ...
##  $ big_stem_bas: num  29.1 25.9 29.5 27 36.2 ...
##  $ can_covr_mea: num  92.2 92.7 94.3 97.9 97.9 ...
##  $ can_heig_mea: num  22.6 22.9 24.3 18.9 26 26.8 15.5 29 27.7 21.2 ...
##  $ ELT         : chr  "do" "do" "do" "do" ...
```

Membaca data dan mengecek bentuk data

Data struktur vegetasi (`veg_str`) berisi 6 variabel:

Kolom	Maksud
X	Titik hitung vegetasi dan burung (metode *point count*)
all_stem_den	Jumlah batang pohon per hektar
big_stem_bas	Luas basal pohon (m^2/ha) dengan diameter min 8 cm pada ketinggian dada (DBH)
can_cover_mea	Persen tutupan kanopi
can_heig_mea	Rataan tinggi kanopi (m)
ELT	*Ecological Land Type* atau tipe habitat

Mengubah bentuk data sesuai tujuan analisis

Nama kolom perlu diubah agar lebih informatif sesuai kebutuhan.

```
colnames(veg_str)<-c("site", "stemDensity", "bigStemBasalArea", "canopyCover", "canopyHeight", "
```

Tipe data karakter dalam `veg_str` juga perlu diganti menjadi faktor.

```
veg_str<-veg_str %>% mutate(site=as.factor(site),
                            landtype=as.factor(landtype))
```

Cek apakah keseluruhan data sudah terganti

```
str(veg_str)
```

```
## 'data.frame':    210 obs. of  6 variables:
##  $ site            : Factor w/ 210 levels "1","2","3","4",..: 1 2 3 4 5 6 7 8 9 10 ...
##  $ stemDensity     : num  11017 3200 7975 15294 17181 ...
##  $ bigStemBasalArea: num  29.1 25.9 29.5 27 36.2 ...
##  $ canopyCover     : num  92.2 92.7 94.3 97.9 97.9 ...
##  $ canopyHeight    : num  22.6 22.9 24.3 18.9 26 26.8 15.5 29 27.7 21.2 ...
##  $ landtype        : Factor w/ 3 levels "dm","do","wm": 2 2 2 2 1 3 2 1 3 2 ...
```

Mengubah bentuk data sesuai tujuan analisis

Beberapa fungsi dalam R membutuhkan bentuk data tertentu.

Untuk fungsi `specnumber()` yang akan digunakan, hanya boleh ada kolom spesies sementara baris menunjukkan lokasi *site*.

```
# Ambil semua data bird kecuali kolom pertama & taruh ke data baru
bird2<-bird[,-1]

# Ganti nama baris data baru  dengan kolom pertama data lama
rownames(bird2)<-bird[,1]
```

Lihat bentuk data baru dan bandingkan dengan yang lama

```
bird2[1:3,1:3]
```

```
##   ACFL AMCR AMGO
## 1    0    0    1
## 2    9    0    0
## 3    4    0    0
```

```
bird[1:3,1:3]
```

```
##   X ACFL AMCR
## 1 1    0    0
## 2 2    9    0
## 3 3    4    0
```

Mengubah bentuk data sesuai tujuan analisis

Data komposisi vegetasi berstruktur sama dengan burung, dan kita ingin mengubahnya agar bisa digunakan sebagai input `specnumber()`.

Ketika set kode yang sama perlu diulangi, kita bisa membuat fungsi

Kode yang ingin diulangi:

```
# Ambil semua data bird kecuali
# kolom pertama & taruh ke data baru
bird2<-bird[,-1]

# Ganti nama baris data baru dengan
# kolom pertama data lama
rownames(bird2)<-bird[,1]
```

Bentuk fungsi dari kode tersebut:

```
ubahData<-function(x)
  {
  # Ambil semua data bird kecuali
  # kolom pertama & taruh ke data baru
  d<-x[,-1]
  # Ganti nama baris data baru dengan
  # kolom pertama data lama
  rownames(d)<-x[,1]

  # Meminta fungsi menampilkan hasil
  return(d)
}
```

Mengubah bentuk data sesuai tujuan analisis

Dengan demikian kita memiliki fungsi `ubahData()` untuk melakukan serangkaian proses transformasi data untuk input yang berbeda.

```
veg2<-ubahData(veg)
```

Dengan demikian kita punya dua data yang siap dihitung jumlah spesies per titik hitung.

```
# Hitung jumlah spesies per titik hitung
birdSp<-specnumber(bird2)
vegSp<-specnumber(veg2)
```

```
head(birdSp)
```

```
##  1  2  3  4  5  6
## 10 15 14 16 15 16
```

```
head(vegSp)
```

```
##  1  2  3  4  5  6
## 16  8 18 18 15 16
```

Analisis data dengan eksplorasi (statistik deskriptif)

Pertama, kita gabung dulu data jumlah spesies ke data struktur vegetasi.

```
veg_str$birdRichness<-birdSp
veg_str$treeRichness<-vegSp
str(veg_str)
```

```
## 'data.frame':    210 obs. of  8 variables:
##  $ site            : Factor w/ 210 levels "1","2","3","4",..: 1 2 3 4 5 6 7 8 9 10 ...
##  $ stemDensity     : num  11017 3200 7975 15294 17181 ...
##  $ bigStemBasalArea: num  29.1 25.9 29.5 27 36.2 ...
##  $ canopyCover     : num  92.2 92.7 94.3 97.9 97.9 ...
##  $ canopyHeight    : num  22.6 22.9 24.3 18.9 26 26.8 15.5 29 27.7 21.2 ...
##  $ landtype        : Factor w/ 3 levels "dm","do","wm": 2 2 2 2 1 3 2 1 3 2 ...
##  $ birdRichness    : int  10 15 14 16 15 16 22 18 17 14 ...
##  $ treeRichness    : int  16 8 18 18 15 16 28 16 14 24 ...
```

Analisis data dengan eksplorasi (statistik deskriptif)

Lalu cek tren yang menjadi pertanyaan penelitian.

Contoh: bagaimana perbedaan kepadatan pohon per hektar di jenis habitat yang berbeda?

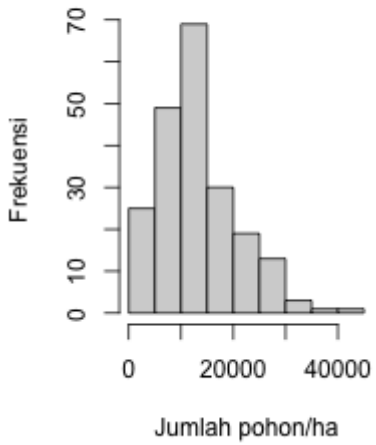
```
veg_str %>% group_by(landtype) %>% summarise(count=n(),
                                             mean=mean(stemDensity),
                                             stdev=sd(stemDensity))
```

```
## # A tibble: 3 × 4
##   landtype count   mean stdev
##   <fct>    <int>  <dbl> <dbl>
## 1 dm          59 11774. 6231.
## 2 do          96 14895. 7976.
## 3 wm          55 12420. 6438.
```

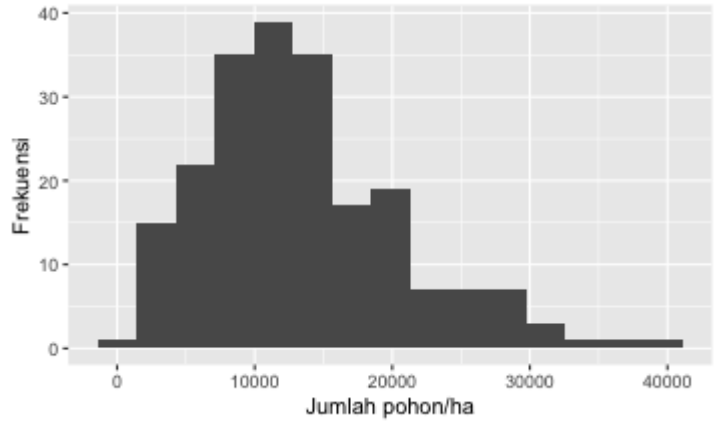
Analisis data dengan eksplorasi (statistik deskriptif)

Lebih mudah melakukan eksplorasi dengan membuat visualisasi data dengan `graphics` (`base R`, kiri) atau `ggplot2` (kanan):

```
hist(veg_str$stemDensity,
     main=NULL,
     xlab="Jumlah pohon/ha",
     ylab="Frekuensi")
```



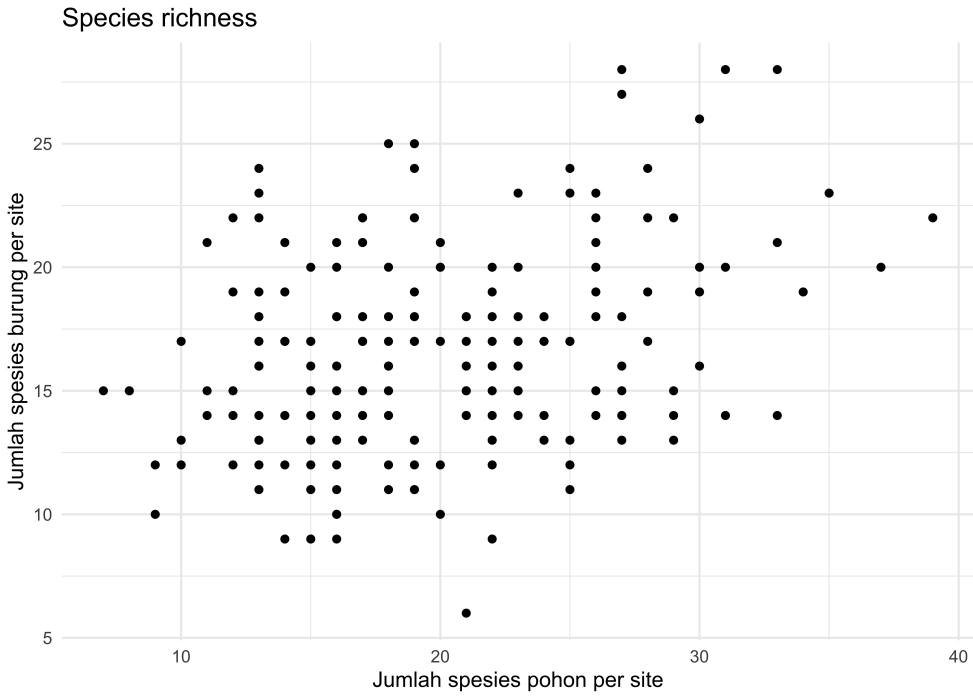

```
ggplot(veg_str)+
  geom_histogram(aes(x=stemDensity), bins=15)
  labs(x="Jumlah pohon/ha",
       y="Frekuensi")
```



Analisis data dengan eksplorasi (statistik deskriptif)

Kita bisa mencoba melihat hubungan antara dua variabel sebagai berikut.

```
ggplot(veg_str)+
  geom_point(aes(x=treeRichness,
                 y=birdRichness))+
    labs(
      x = "Jumlah spesies pohon per site",
      y = "Jumlah spesies burung per site",
      title = "Species richness")+
  theme_minimal()
```



Seputar visualisasi data dengan ggplot, bisa dibaca di: https://r4ds.had.co.nz/data-visualisation.html

Uji Statistik

1. Menentukan hipotesis nol dan alternatif
    - biasanya hipotesis nol adalah "*tidak ada efek*", dan alternatifnya adalah "*ada efek*"
    - perlu diperhatikan bahwa teori bisa memberikan hipotesis nol yang berbeda. Contoh: gradien hubungan keanekaragaman adalah 3/4 sehingga hipotesis nolnya adalah gradien = 3/4 dan alternatifnya adalah negasinya.
2. Menghitung besaran efek
    - perbedaan antar rataan
    - besar gradien hubungan antara dua variabel
    - perbedaan antara dua perlakuan
    - dll (sangat tergantung studi)

Uji Statistik

1. Ubah besaran efek sehingga kita bisa membandingkannya ke suatu distribusi yang diharapkan
    - Kita ingin tahu probabilitas suatu besaran efek sama atau mungkin lebih besar dari yang kita amati, mengasumsikan hipotesis nol benar.
    - Biasanya kita tidak bisa langsung menghitung ini dari besaran efek; di sinilah transformasi matematis diperlukan untuk membuat besaran efek yang kita amati dapat dibandingkan ke suatu distribusi besaran efek (t, F, chi-square, dll.)
2. Tanyakan peluang mengamati efek sebesar atau lebih besar dari yang kita amati, mengasumsikan hipotesis nol benar (P value).
    - Batasan yang paling umum adalah 0.05
3. Pikirkan apa artinya.

Uji Statistik (Uji Korelasi)

Ketika kita ingin melihat hubungan korelasi antara dua variabel, langkah 2-4 bisa kita lakukan dengan fungsi `cor.test()`:

```
cor.test(veg_str$birdRichness,veg_str$treeRichness)
```

```
##
##     Pearson's product-moment correlation
##
## data:  veg_str$birdRichness and veg_str$treeRichness
## t = 4.3821, df = 208, p-value = 1.864e-05
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.1616938 0.4099734
## sample estimates:
##       cor
## 0.2907198
```

Uji Statistik (Uji Korelasi)

Perhatikan bahwa untuk melakukan uji korelasi, data yang diuji harus memenuhi beberapa asumsi:

- Distribusi variabel x dan y mendekati distribusi normal
- Variansi y tetap sama sepanjang nilai x dan sebaliknya
- Hubungan antara x dan y mendekati garis lurus (bukan kurva)

Lebih banyak lagi tentang analisis korelasi di http://research.sbcs.qmul.ac.uk/r.knell/Intro_biostats_R/correlation-analysis.html

Uji Statistik (Regresi linear)

Kita juga bisa melakukan regresi linear, atau memodelkan hubungan antara variabel x dan y agar mengikuti persamaan berikut:

$$y = a + bx$$

di mana y adalah variabel dependen, x variabel independen, a adalah konstanta yang mendefiniskan nilai y ketika x = 0, dan b adalah gradien garis hubungan antara x dan y.

Asumsi yang harus dipenuhi data:

- Independen
- Nilai *error* (perbedaan antara nilai observasi dengan nilai prediksi dari garis) mengikuti distribusi normal
- Variansi y tetap sama sepanjang nilai x dan sebaliknya (*homogeneity principle*)
- Ada hubungan linear antara variabel independen (x) dan variabel dependen (y)
- Variabel independen diukur tanpa error (umumnya hanya perlu dikawatirkan dalam studi alometri ketika ukuran bagian tubuh berkorelasi dengan ukuran tubuh keseluruhan)

Lebih banyak lagi tentang analisis korelasi di http://research.sbcs.qmul.ac.uk/r.knell/Intro_biostats_R/using-linear-regression-to-analyse-tb-trends-in-the-uk.html#assumptions-of-linear-regression

Uji Statistik (Regresi linear)

Mari coba lihat hubungan jumlah spesies pohon dan burung dalam satu habitat:

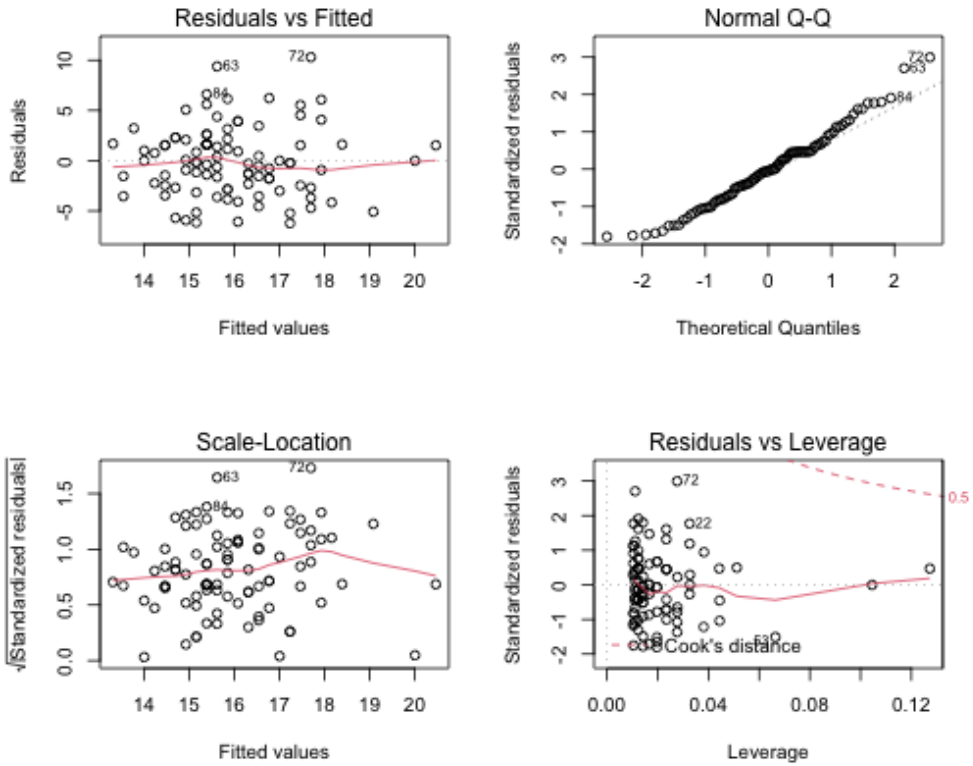
```
m_do<-veg_str %>% filter(landtype=="do") %>% lm(birdRichness ~ treeRichness,.)
summary(m_do)
```

```
##
## Call:
## lm(formula = birdRichness ~ treeRichness, data = .)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.2365 -2.5246 -0.1581  1.6306 10.3017
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  11.46314    1.24565   9.203 9.12e-15 ***
## treeRichness  0.23093    0.06118   3.774 0.000281 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.488 on 94 degrees of freedom
## Multiple R-squared:  0.1316,    Adjusted R-squared:  0.1224
```

Uji Statistik (Regresi linear)

Untuk mengetahui apakah model linear yang kita buat memenuhi asumsi, kita dapat melihat plot diagnostik `lm()` dalam R.

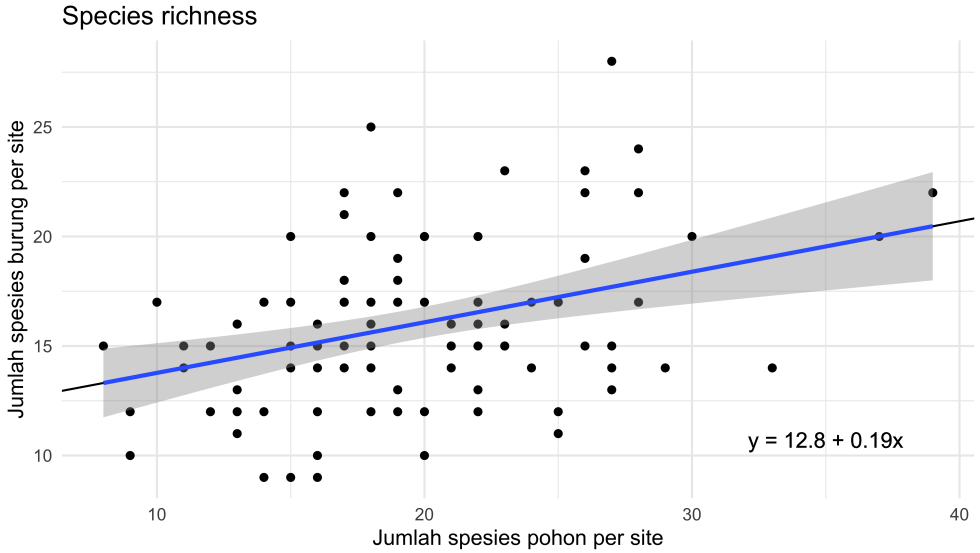
```
par(mfrow=c(2,2))
plot(m_do)
```



Uji Statistik (Regresi linear)

Ketika kita sudah yakin bahwa model regresi linear kita menjelaskan dengan baik hubungan antara dua variabel, kita perlu memikirkan apa artinya.

```
veg_str %>%
  filter(landtype=="do") %>%
  ggplot()+
  geom_point(aes(x=treeRichness,y=birdRichnes
    labs(x = "Jumlah spesies pohon per site",
       y = "Jumlah spesies burung per site",
       title = "Species richness")+
  geom_abline(slope=0.23093, intercept=11.463
  geom_smooth(aes(x=treeRichness,y=birdRichne
  geom_text(aes(x=35,y=10), label=paste("y =
  theme_minimal()
```



Misal, berapa banyak pertambahan spesies burung per pohon?

Uji Statistik (Regresi linear)

Berdasarkan hasil regresi, gradien garis adalah 0.23, jadi ada 1 spesies burung setiap 0.23 spesies pohon.

Seberapa meyakinkan angka ini? Fungsi `confint()` memberikan kita 95% selang kepercayaan untuk estimasi angka a dan b.

```
confint(m_do)
```

```
##                   2.5 %     97.5 %
## (Intercept)  8.9898711 13.9364139
## treeRichness 0.1094536  0.3524132
```

Berdasarkan angka ini kita 95% yakin bahwa gradien yang sesungguhnya ada di antara angka 0.11 dan 0.35.

Uji Statistik (Regresi linear)

Berapa banyak spesies burung yang kita harapkan dari 20 hingga 25 spesies pohon misalnya?

```
a=11.46314
b=0.23093
x=c(20:25)

y<-a+b*x
y
```

```
## [1] 16.08174 16.31267 16.54360 16.77453 17.00546 17.23639
```

Hanya 16-17 spesies burung.

Uji Statistik

Uji statistik "hanya" alat, kita perlu memikirkan apa artinya dan bagaimana hipotesis yang kita miliki menentukan uji statistik yang kita perlukan.

Misal, apakah cukup menggunakan jumlah spesies pohon untuk memprediksi kekayaan burung? Apakah tidak perlu variabel yang lain?

Banyak konsultasi dengan dosen pembimbing mengenai variabel yang perlu dieksplorasi dan diuji.

Latihan: Apakah ada hubungan antara struktur vegetasi dengan kekayaan spesies burung?

1. Buat R `Project` dan struktur folder untuk yang sesuai

2. Unduh data di https://datadryad.org/stash/dataset/doi:10.5061/dryad.k48h616 (klik '*Download dataset*'), ekstrak datanya, dan letakkan di folder input kalian sebagai data mentah

3. Baca data yang sesuai ke dalam R (lihat `README.txt`) dan lakukan cleaning data seperlu kalian untuk menjawab pertanyaan hubungan antara struktur vegetasi dan kekayaan spesies burung (pilih salah satu parameter struktur vegetasi)

Selamat mencoba!

Slide ini dapat diakses di
https://github.com/sagitaninta/hardSkillNymphaea

Slides created via the R package **xaringan**.

The chakra comes from remark.js, **knitr**, and R Markdown.