Iowa Social Science Research Center
2020-21 Workshop series

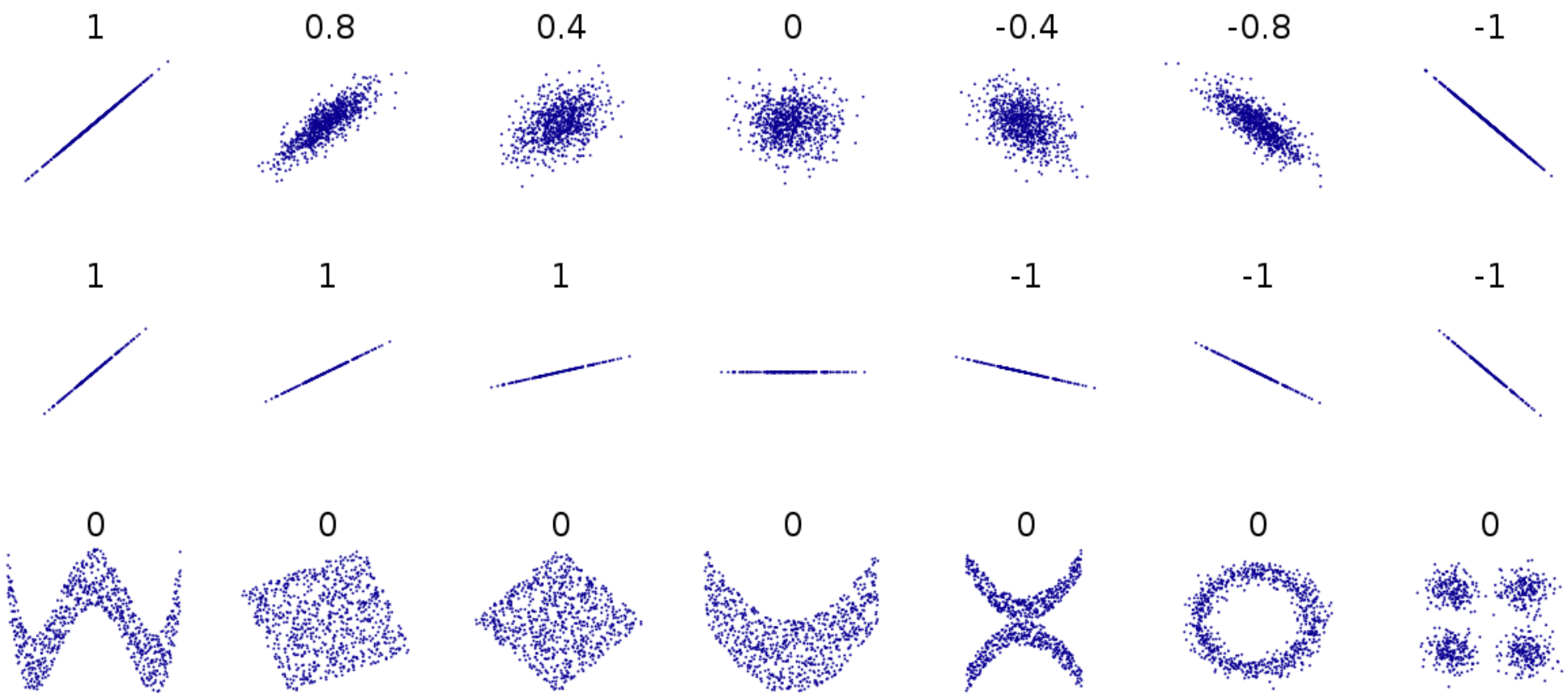
THE UNIVERSITY OF IOWA

Advanced Spatial analysis with R

Seungwon Kim

Geographical and Sustainability Sciences

Correlation



Regression

- A set of statistical processes for estimating the relationships between a dependent variable and one or more independent variables

- Goals
  - Identify the statistically significant predictors
    - Education, Epidemiology,…
  - Prediction

Ordinary Least Squares Regression (OLS)

- y_i is the **response variable**
  - a.k.a. Dependent variable
- x_i is the **explanatory variable** or **predictor variable**
  - a.k.a. Independent variable

- We are using X to explain some or most of the variability of y, in particular, the non-random part of y's variability

Steps for linear regression

- Response variable
  - Type of variable
    - **Continuous: Simple or multiple regression**
    - Count data: Poisson regression or negative binomial regression
    - Binary: Logistic/Probit regression
  - Independent?
    - Study design
    - Ex) School, classroom, students
- Linear relationship between response and predictor(s)
  - Scatter plot

Steps for linear regression (cont.)

- Predictor
  - Check multicollinearity
    - Correlation matrix
    - Variance Inflation Factor (VIF)
- Error
  - Normality
    - Q-Q plot
    - Shapiro-Wilk test
      - H_0: The population is normally distributed
  - Heteroscedasticity
    - Plot (studentized) residuals vs. predicted y
    - Breusch-Pagan test
      - H_0: Homoskedasticity
  - Autocorrelation
    - Plot
    - Burbin-Watson test

The first law of geography

- Waldo Tobler
  - “Everything is related to everything else, but near things are more related than distant things.”

- Spatial dependence
  - The co-variation of properties within geographic space
  - Weather map (Kriging)
  - Spatial autocorrelation

Back to regression analysis

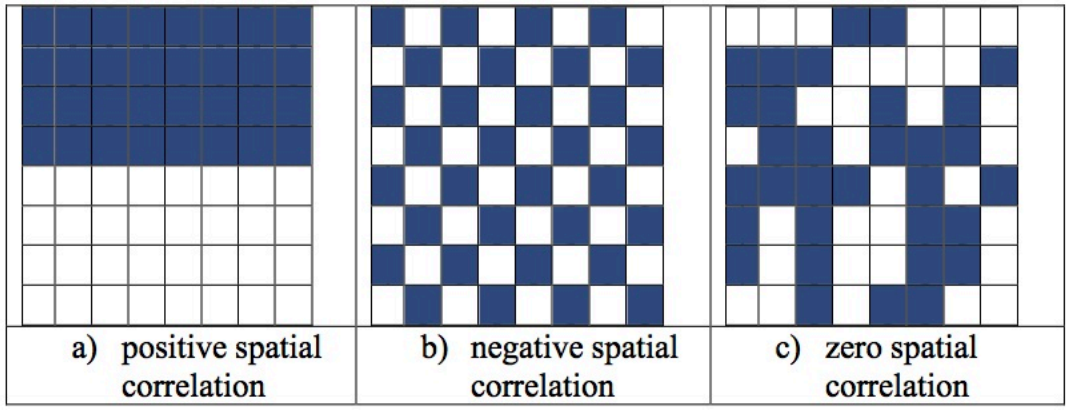
- Response variable
  - Type of variable
    - **Continuous: Simple or multiple regression**
    - Count data: Poisson regression or negative binomial regression
    - Binary: Logistic regression
  - Independent? (Spatially independent?)
    - Study design
    - Ex) School, classroom, students

Back to regression analysis

- Error
  - Normality
    - Q-Q plot
    - Shapiro-Wilk test
  - Heteroscedasticity
    - Plot (studentized) residuals vs. predicted y
    - Breusch-Pagan test
  - Autocorrelation (Spatial autocorrelation)
    - Plot
    - Burbin-Watson test

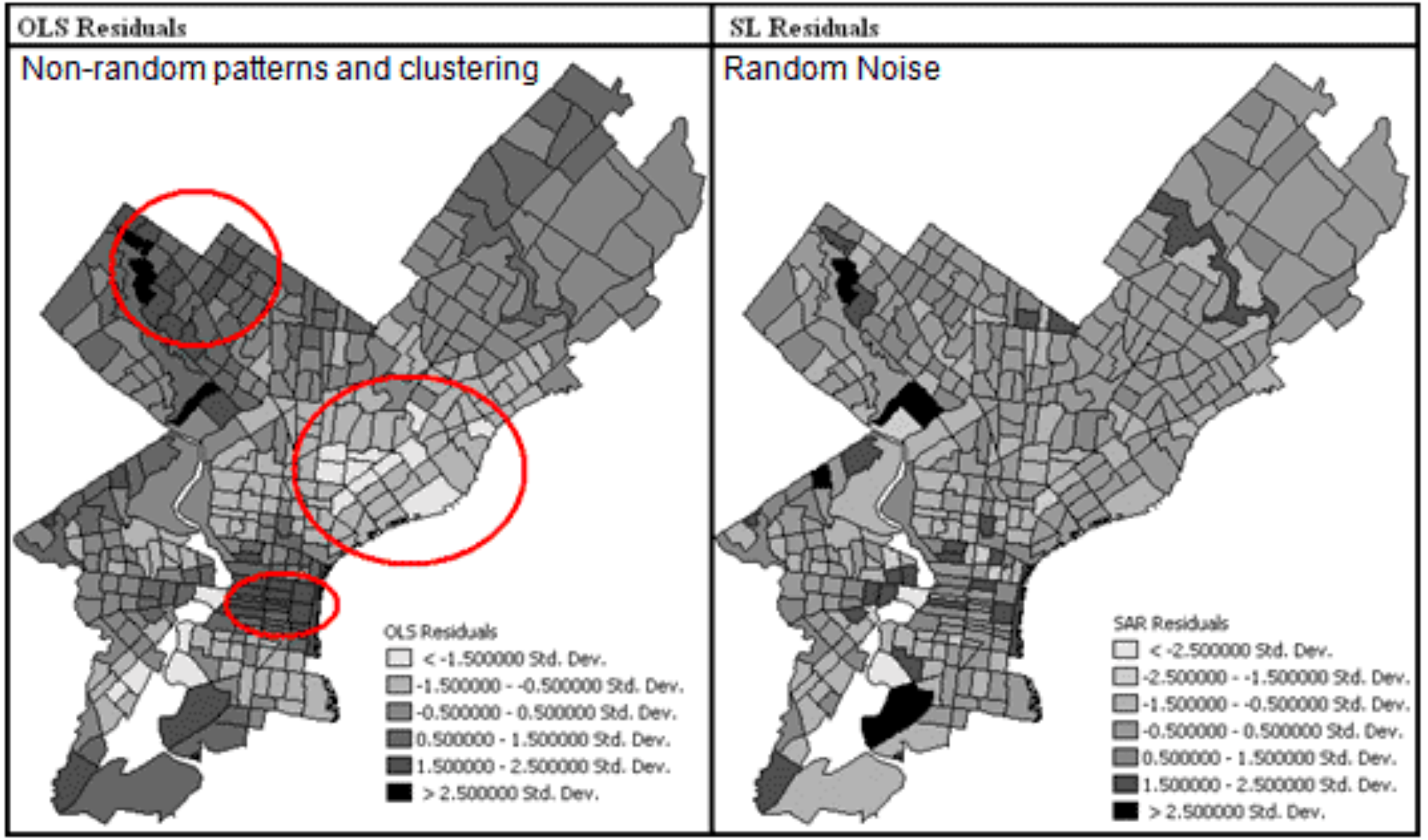
Check Spatial dependence

- Moran's I
  - A measure of spatial autocorrelation



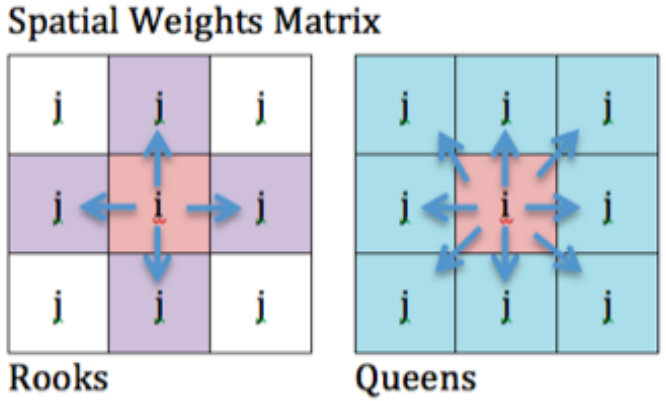
  - A summary statistic with p-value from bootstrapping

Check Spatial dependence – residual map



Spatial Weight matrix

- The spatial relationships that exist among the features in your dataset
- NxN matrix
- Two ways to determine neighbors
  - Rooks case
  - Queens case



Spatial regression models

- Spatial Lag model
  - Assumes that dependencies exist directly among the levels of the response variable
  - Used when we know the structure of spatial dependence
  - $Y = \rho WY + X\beta + \varepsilon$

- Spatial Error model
  - Error term has dependence
  - Used when structure of dependence is unknown
  - $Y = X\beta + \lambda W\varepsilon + \xi$

Data Structures

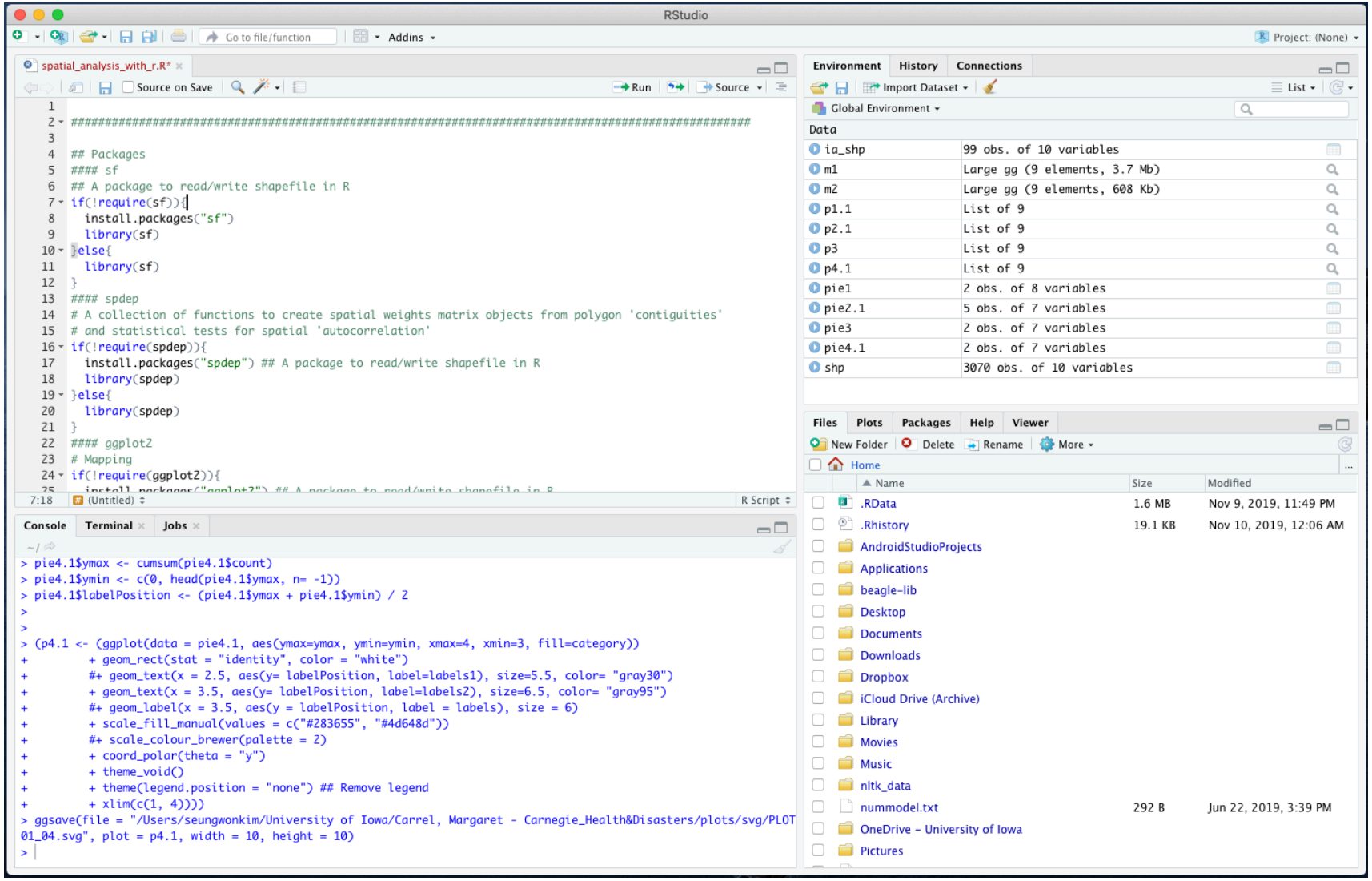
- Vectors
  - A 1-dimensional object that consists of indexed elements of the same data type
  - Numeric / Character / Date / Logical / Factors
  - X <- c(1, 2, 3, 4, …)
  - X[i]

- Matrix
  - A 2-dimensional object that consists of indexed elements of the same data type
  - X[i, j]

Data Structures

- Data frame
  - A 2-dimensional object that consists of indexed elements
  - Elements within a given column are of the same type, but types may differ between columns.
  - x[i, j]
  - x[i,] ## Extract a row
  - x[, j] ## Extract a column

R Studio


RStudio
spatial_analysis_with_r.R*
Source on Save
Run
Source
Environment
History
Connections
Import Dataset
List
Global Environment
Data
ia_shp 99 obs. of 10 variables
m1 Large gg (9 elements, 3.7 Mb)
m2 Large gg (9 elements, 608 Kb)
p1.1 List of 9
p2.1 List of 9
p3 List of 9
p4.1 List of 9
pie1 2 obs. of 8 variables
pie2.1 5 obs. of 7 variables
pie3 2 obs. of 7 variables
pie4.1 2 obs. of 7 variables
shp 3070 obs. of 10 variables
Files
Plots
Packages
Help
Viewer
New Folder
Delete
Rename
More
Home
Name
Size
Modified
.RData 1.6 MB Nov 9, 2019, 11:49 PM
.Rhistory 19.1 KB Nov 10, 2019, 12:06 AM
AndroidStudioProjects
Applications
beagle-lib
Desktop
Documents
Downloads
Dropbox
iCloud Drive (Archive)
Library
Movies
Music
nltk_data
nummodel.txt 292 B Jun 22, 2019, 3:39 PM
OneDrive - University of Iowa
Pictures
Console
Terminal
Jobs
R Script