Creative Data Mining

Lecture 02: Intro to RStudio and Clustering
29 February 2016

Danielle Griego, griegod@ethz.ch
Matthias Standfest, standfest@arch.ethz.ch

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

DARCH iA Chair of Information Architecture

What we'll cover today:

- HW1: Share what you have learned about urban data mining (20 min)
- Break (10 min)
- R tutorial (60 min)

Homework_1

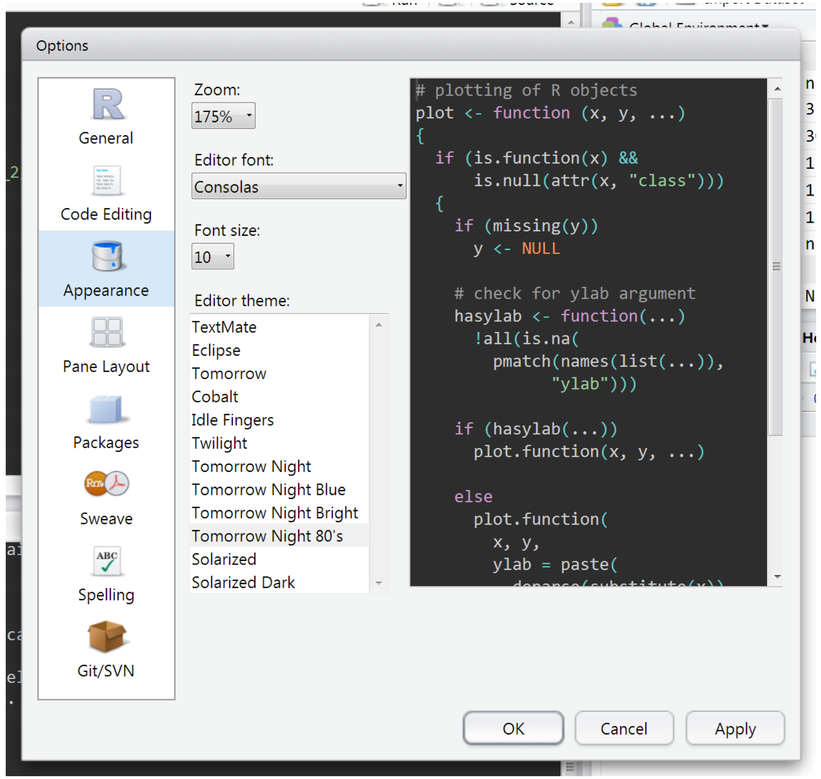
1. Install R from http://cran.r-project.org/

2. Install RStudio from http://www.rstudio.com/products/rstudio/download/

3. Research other examples of urban data mining and make 3 slides about the most interesting project/application/resarch group(s) that you find. Will be presented at the beginning of next lecture.

CREATIVE DATA MINING
Intuitively Analysing Design Ideas

ADJUST COLORS



ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

DARCH iA Chair of Information Architecture

CREATIVE DATA MINING

Intuitively Analysing Design Ideas

OVERVIEW

lecture_2.R

```r
#
# CREATIVE DATA MINING - FS 2016
# LECTURE 2 - R101
# Matthias Standfest
# Danielle Griego

## OVERVIEW

# RESOURCES
# keyboard shortcuts: https://support.rstudio.com/hc/en-us/articles/200711853-Keyboard-Shortcuts
# help() function

# NUMBERS
a <- 2 #store a number
(b <- 5) #brackets for writing to console

(a*b*pi) #simple multiplication with predefined variable

# VECTORS
(vector_a <- c(2,3,4,5)) #save a list of numbers
(vector_b <- 5:8) #shortcut for number sequence

(vector_a * vector_b) #multiplying lists

# MATRICES
(matrix_a <- matrix(1, nrow=4, ncol=4)) #create 4 x 4 matrix filled with 1
(matrix_b <- matrix(1:16, nrow=4, ncol=4)) #fill matrix with number sequence
(matrix_c <- matrix(1:16, nrow=4, ncol=4, byrow=TRUE)) #fill by rows not by columns

matrix_b1<- t(matrix_b) #transposing also exchanges rows and columns

(matrix_a *3) #multiplying matrix with scalar
(matrix_b * matrix_c) #multiplying matrices elementwise
test <- matrix_b %*% matrix_c #dot product of two matrices

# PLOTTING
plot(vector_a, type="o", col="blue") #simple plot of vector
help(plot) #search plot elements, change plot type, col, etc.

#install.library(fields)
#library(fields) #import library need to import libraries for certain functions

image.plot(matrix((data=matrix_a), ncol=4, nrow=4)) #plot matrix_a, compare plot with numeric values of matrix_a
image.plot(matrix((data=matrix_b), ncol=4, nrow=4)) #plot matrix_b, compare plot with numeric values of matrix_b
image.plot(matrix((data=matrix_b1), ncol=4, nrow=4)) #plot matrix_b1, compare plot with numeric values of matrix_b1
image.plot(matrix((data=matrix_c), ncol=4, nrow=4)) #plot matrix_c, compare with matrix_b and matrix_b1

# Note that you can transpose data in a matrix using byrow=TRUE when creating the matrix, or by t(matrix_b)

```

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

DARCH iA Chair of Information Architecture

IRIS DATA SET (1936)

en.wikipedia.org/wiki/Iris_flower_data_set

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

DARCH iA Chair of Information Architecture

VIEW PREINSTALLED DATA FRAME

lecture_2_00.R

```
#
# CREATIVE DATA MINING - FS 2016
# LECTURE 2 - R101
# Matthias Standfest
# Danielle Griego

## EXPLORING PREINSTALLED DATA

# 1. set your working directory
# or use Ctl+Shift+H and select
setwd("~/Dropbox/00_Work/01_Teaching/Creative Data Mining/001_FS15-DataMining/lecture_2/RSCRIPTS_FS16/")

# 2. print the preinstalled data set
print(iris)

# 3. print first rows
head(iris)

# 4. print structural information
str(iris)
colnames(iris)

# 5. print summary
summary(iris)

```

SAVE AND READ DATA

lecture_2_01.R

```r
#
# CREATIVE DATA MINING - FS 2016
# LECTURE 2 - R101
# Matthias Standfest
# Danielle Griego

## READING AND WRITING CSV FILES TO A FOLDER

# 1. set your working directory
setwd("~/Dropbox/00_Work/01_Teaching/Creative Data Mining/001_FS15-DataMining/lecture_2/RSCRIPTS_FS16/")

# 2. use data
myData <- iris

# 3. write data in file, check that it's in your working directory
write.table(myData, file = "myFlowerData.csv", sep = ";", col.names = NA, qmethod = "double")

# 4. read data back into R
myFile <- read.table("myFlowerData.csv", header = TRUE, sep = ";", row.names = 1)

# 5. write data in R file (smaller and better structured), check that it's in your working directory
saveRDS(myData, file="myFlowerDataUnreadableButBetter.RData")

# 6. get data from R file
myOtherFile <- readRDS(file="myFlowerDataUnreadableButBetter.RData")

```

USE ONLY PARTS OF THE DATA

lecture_2_02.R

```r
#
# CREATIVE DATA MINING - FS 2016
# LECTURE 2 - R101
# Matthias Standfest
# Danielle Griego

## ACCESSING SPECIFIC ROWS AND COLUMNS FROM A MATRIX

# 1. set your working directory
setwd("~/Dropbox/00_Work/01_Teaching/Creative Data Mining/001_FS15-DataMining/lecture_2/RSCRIPTS_FS16/")

# 2. use data partially
(iris[1:3,]) # rows 1 TO 3

(iris[1:3,3:5]) # rows 1 TO 3 and columns 3 TO 5

(iris[1:3,c(1,5)]) # rows 1 TO 3 and columns 1 AND 5

# 3. Question: create a matrix from the iris dataset without categorical data (column 5)
myData <- iris[,1:4]

```

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

DARCH iA Chair of Information Architecture

DATAFRAME AND MATRIX

lecture_2_03.R

```r
#
# CREATIVE DATA MINING - FS 2016
# LECTURE 2 - R101
# Matthias Standfest
# Danielle Griego

## MATRIX VERSUS DATA FRAMES?

# 1. set your working directory
setwd("~/Dropbox/00_Work/01_Teaching/Creative Data Mining/001_FS15-DataMining/lecture_2/RSCRIPTS_FS16/")

# 2. WHAT IS A DATA FRAME?
## answer <- a data frame is a list of vectors with equal length, thus a matrix with mixed types

# 2.1. WHAT ARE TYPES?
# answer <- we have numerics (numbers), characters (text) and logicals (TRUE or FALSE)
numeric_vector <- c(1,3.14,23,42)
character_vector <- c("start", "pie", "the all seing eye", "the answer to everything")
logical_vector <- c(TRUE, TRUE, FALSE, TRUE)

# try decomposing the vector structure to understand how it works, for example remove the c, a generic function which combines its
logical_vector <- (TRUE, TRUE, FALSE, TRUE)

# combine vectors to a data.frame
my_first_dataframe <- data.frame(numeric_vector, character_vector, logical_vector)

# 4. WHAT IS A MATRIX?
# answer <- a matrix is a list of vectors with equal length of the SAME type
myMatrix <- matrix(1:20, nrow = 5, ncol = 4)
# label the matrix rows and columns
dimnames(myMatrix) <- list(c("a","b","c","d","e"), character_vector)
# use those labels
myMatrix[,c("start","the answer to everything")] #to identify columns
myMatrix["a",] #to identify a single row (or more)

# 5. labelling a data frame
names(my_first_dataframe) <- c("values","nerdstuff","BOOL")
my_first_dataframe["values"] # access like a matrix
my_first_dataframe$values # shortcut to a vector

```

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

DARCH iA Chair of Information Architecture

ADDITIONAL STUFF

lecture_2_04.R

```r
#
# CREATIVE DATA MINING - FS 2016
# LECTURE 2 - R101
# Matthias Standfest
# Danielle Griego

## YOU CAN ALSO ORGANIZE DATA IN LISTS

# 1. set your working directory
setwd("~/Dropbox/00_Work/01_Teaching/Creative Data Mining/001_FS15-DataMining/lecture_2/RSCRIPTS_FS16/")

# 2. LISTS
# additional type <- a list is just an ordered collection of all objects under one name
myList <- list(letters = c("a","b","c"),
               numbers = 1:50,
               datatable = matrix(1:2, nrow=2, ncol=2),
               answer = 42,
               innerlist = list(sequence2=1:25, sequence3=26:35)) # you can use linebreaks to make your code more readable

myList[2] #access element by index
myList$numbers #access element by name
myList[["numbers"]] #access element via double bracket convention
myList$innerlist$sequence3 #access nested element
myList$innerlist #access nested element

# 3. additional functions
length(myList) # number of elements
class(myList$numbers) #class of element
class(myList$letters) # now try this for another variable
cbind(1,2,3) #instead of c, binding in matrix form
rbind(1,2,3) #matrix bind row wise
ls() # list all current objects, notice they are also in the upper righthand object pane

```

Schedule

Creative Data Mining
Intuitively Analysing Design Ideas

The goal of this course is to introduce various data mining techniques for design and urban planning applications. Students will learn how to select relevant data sources and collect their own data using a "sensor backpack". Various methods will be applied to a common project to evaluate the predominant influencing factors of the urban environment on our perceptual experiences. A select neighborhood in the city will be used as a case study. Final results will be presented in the last class.

The course will start with an initial overview to data mining and the relevant mathematics as well as an introduction to the programming tool (RStudio). Then students will learn how to use and interpret results from a machine-learning tool to cluster self-made design sketches, which automatically generate qualitative collages. Finally, students will collect data using a "sensor backpack" with environmental sensors such as noise, temperature, illuminance, and air particulates. Students will also generate the data for perceptual quality in this neighborhood through time-stamped and geo-referenced surveys and biofeedback wristbands. Students will be given a work-flow to collect, process, analyze and interpret this data which may be used in their final projects.

Where
HIT H 12

Supervision
Danielle Griego griego@arch.ethz.ch
Matthias Standfest standfest@arch.ethz.ch

22.02.2016	**Course Introduction** Introduce data-mining techniques and case study
29.02.2016	**Introduction to the Environment** Introduction to R Studio and clustering
07.03.2016	**From analog to digital analysis** Use hand-drawn sketched to auto-generated collages
14.03.2016	**Seminar week (No lecture)**
21.03.2016	**Analysis and interpretation I** Evaluate auto-generated collages
28.03.2016	**Holiday (No lecture)**
04.04.2016	**Time-series data analysis and Urban Planning** Introduction to time-series analysis
11.04.2016	**Data collection with sensor backpack** Collect data and introduce workflows
18.04.2016	**Holiday (No lecture)**
25.04.2016	**Analysis and interpretation II** Evaluate sensor backpack data
02.05.2016	**Q&A Feedback Workshop** Finalise semester projects
09.05.2016	**Final iA critique** Combined critique with the other iA courses (14:00 - 16:00)

Requirement: Former knowledge of any digital tool or coding language is most welcome but NOT required. You only need to provide a reasonable amount of motivation and of course a notebook.

* Total 60 h = 2 ECTS
Exercises 40% (documentations)
Final Presentation 40% (Final project)
Attendance 20%

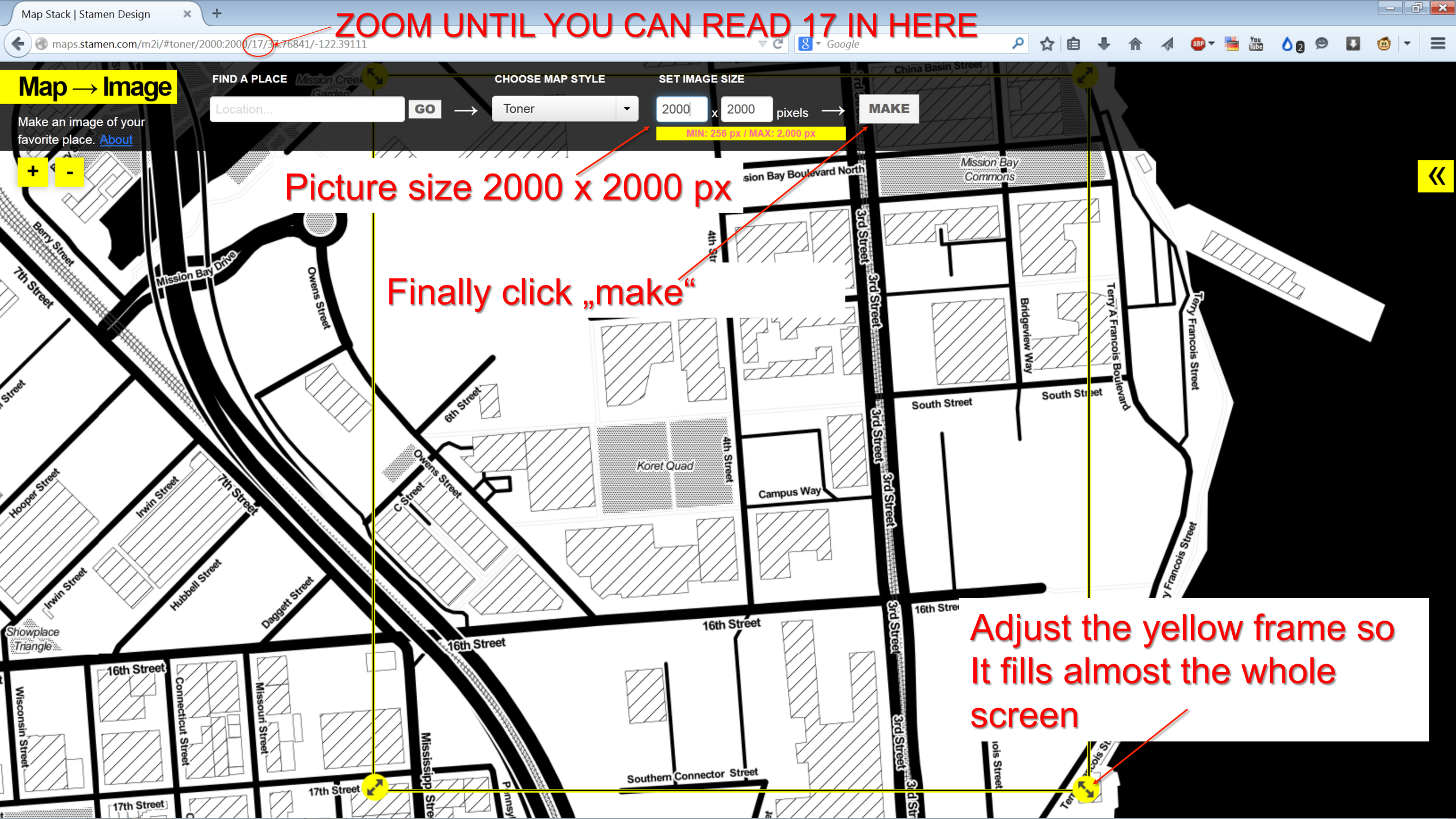
The most recent outline will be found on www.ia.arch.ethz.ch

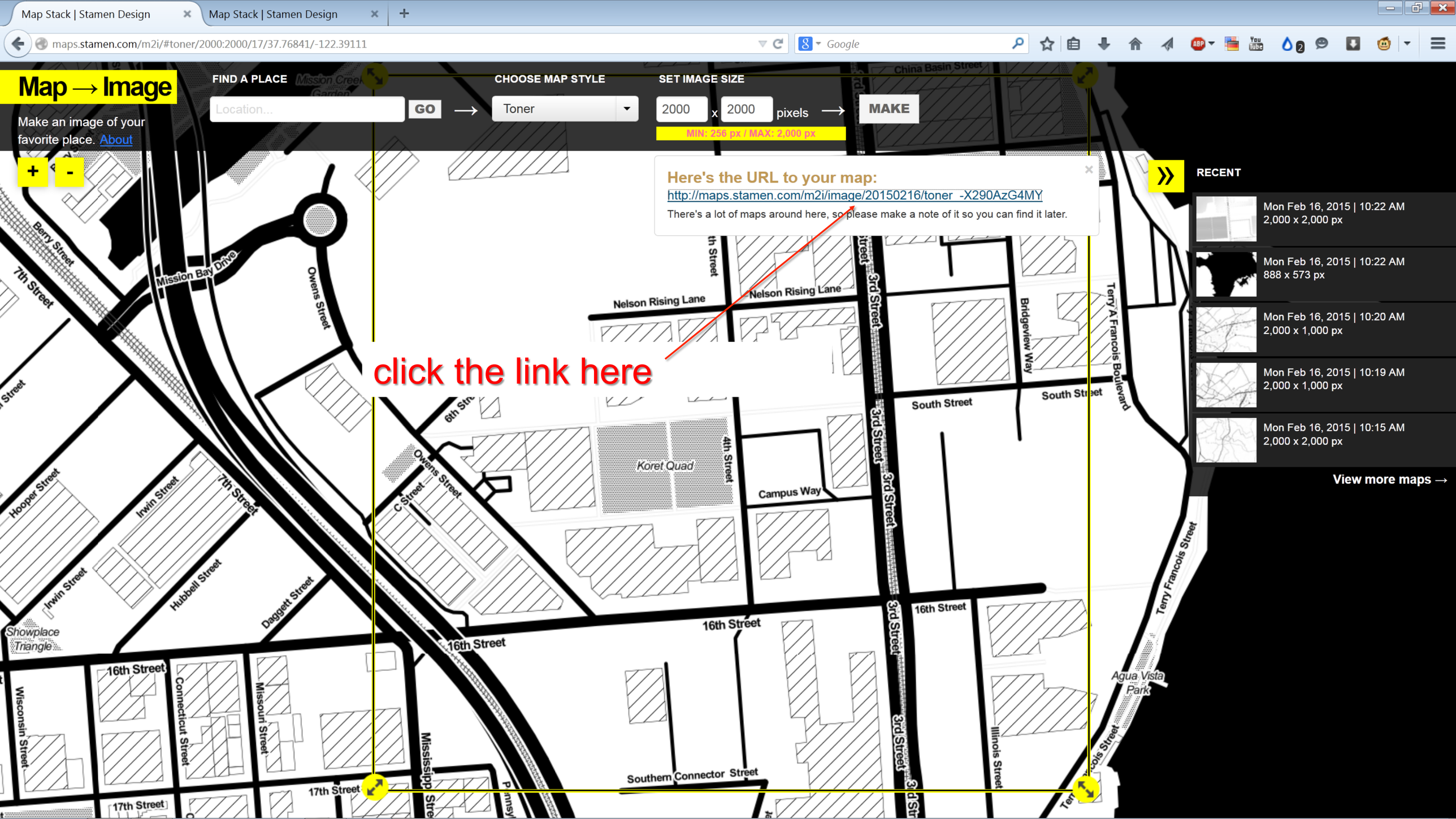
ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

DARCH iA Chair of Information Architecture

Homework:

1. Review the R-tutorials lecture_2_00 through lecture_2_04
2. Complete HW_2
3. Optional: Review how to export images from maps.stamen.com according to the tutorial on the next pages. We will prepare the street maps for the next course exercise, but this is one method to create an image database for clustering analysis in the first block.


Map Stack | Stamen Design
ZOOM UNTIL YOU CAN READ 17 IN HERE
maps.stamen.com/m2i/#toner/2000:2000/17/37.76841/-122.39111
Map → Image
Make an image of your favorite place. About
FIND A PLACE
Location...
GO
CHOOSE MAP STYLE
Toner
SET IMAGE SIZE
2000 x 2000 pixels
MIN: 256 px / MAX: 2,000 px
MAKE
Picture size 2000 x 2000 px
Finally click „make“
Adjust the yellow frame so
It fills almost the whole
screen
China Basin Street
Mission Bay Boulevard North
Mission Bay Commons
Berry Street
7th Street
Mission Bay Drive
Owens Street
6th Street
4th Street
3rd Street
Bridgeview Way
Terry A Francois Boulevard
Terry Francois Street
South Street
Koret Quad
Campus Way
C Street
Hooper Street
Irwin Street
Hubbell Street
Daggett Street
Showplace Triangle
16th Street
Wisconsin Street
Connecticut Street
Missouri Street
Mississippi Street
17th Street
Southern Connector Street

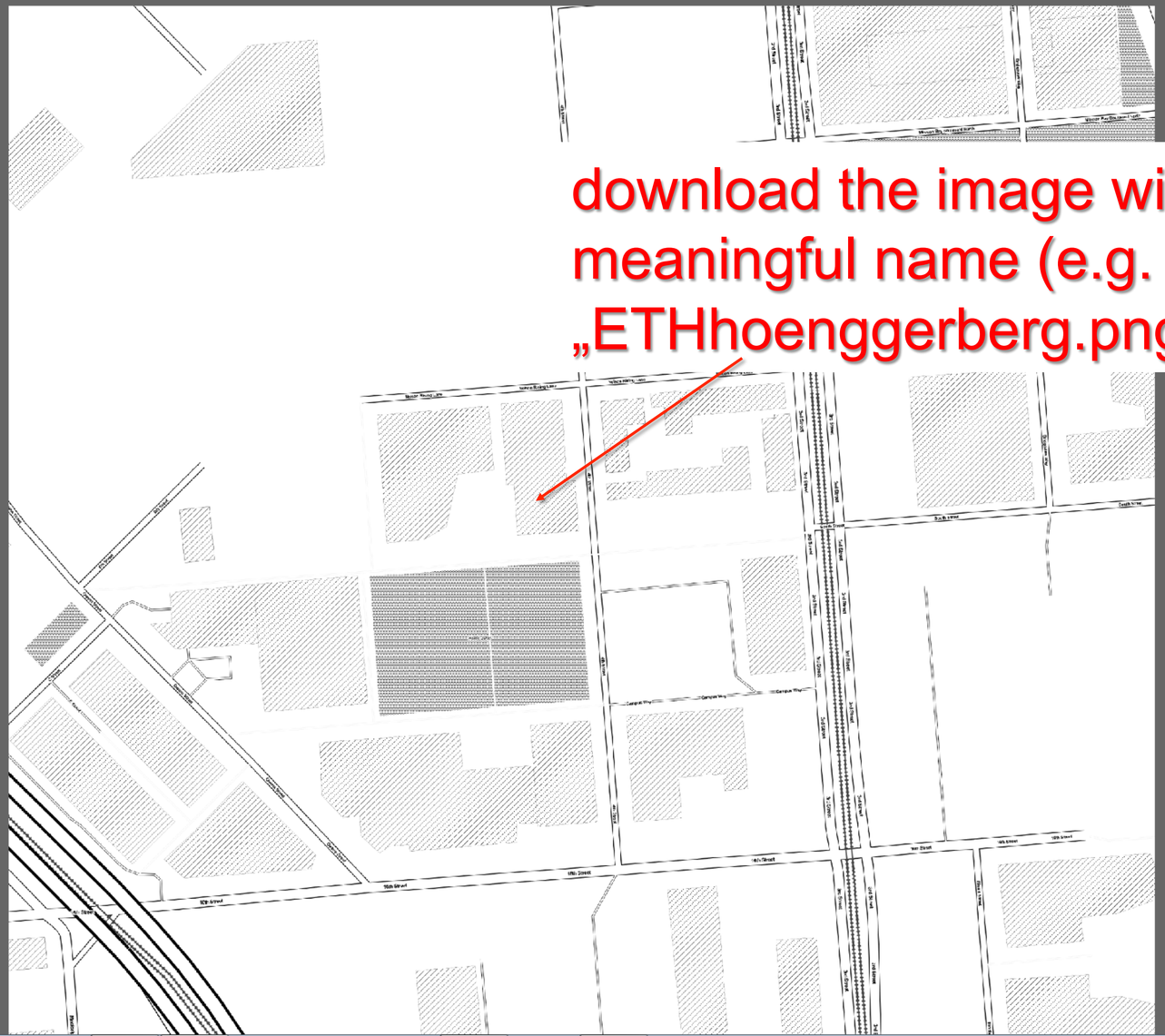

Map Stack | Stamen Design
Map Stack | Stamen Design
maps.stamen.com/m2i/#toner/2000:2000/17/37.76841/-122.39111
Google
Map → Image
Make an image of your favorite place. About
FIND A PLACE
Location...
GO
CHOOSE MAP STYLE
Toner
SET IMAGE SIZE
2000
x
2000
pixels
MIN: 256 px / MAX: 2,000 px
MAKE
Here's the URL to your map:
http://maps.stamen.com/m2i/image/20150216/toner_-X290AzG4MY
There's a lot of maps around here, so please make a note of it so you can find it later.
click the link here
RECENT
Mon Feb 16, 2015 | 10:22 AM
2,000 x 2,000 px
Mon Feb 16, 2015 | 10:22 AM
888 x 573 px
Mon Feb 16, 2015 | 10:20 AM
2,000 x 1,000 px
Mon Feb 16, 2015 | 10:19 AM
2,000 x 1,000 px
Mon Feb 16, 2015 | 10:15 AM
2,000 x 2,000 px
View more maps →
Berry Street
7th Street
Mission Bay Drive
Owens Street
Nelson Rising Lane
3rd Street
Bridgeview Way
Terry A Francois Boulevard
South Street
Koret Quad
4th Street
Campus Way
C Street
Hooper Street
Irwin Street
Hubbell Street
Daggett Street
Showplace Triangle
16th Street
Wisconsin Street
Connecticut Street
Missouri Street
Mississippi Street
Southern Connector Street
17th Street
Terry Francois Street
Agua Vista Park
Illinois Street

Map Stack | Stamen Design

maps.stamen.com/m2i/image/20150216/toner_-X290AzG4MY

Map → Image → 2015.02.16

Make an image of your favorite place About



download the image with a meaningful name (e.g. „ETHhoenggerberg.png“)