Complex Systems & Network Thinking

Pierre-Alexandre Balland

Organizations and Networks

Economic and Organizational structures

ON is about applying network thinking (and complex system thinking) to solve economic and business problems

Today’s objectives

- Real world networks in science and business

Today’s objectives

- Real world networks in science and business
- Network thinking

Today’s objectives

- Real world networks in science and business
- Network thinking
- Link organizations and networks

Today’s objectives

- Real world networks in science and business
- Network thinking
- Link organizations and networks
- Structure of the class (topics, exam, project, …)

On the side

- Discussion on big data
- Do we still need theory when we have big data?
- Data visualization techniques – art or science?

Lab #1

- Discuss project idea (& start forming groups)
- Introduction to R, RStudio and R packages
- First programming attempt

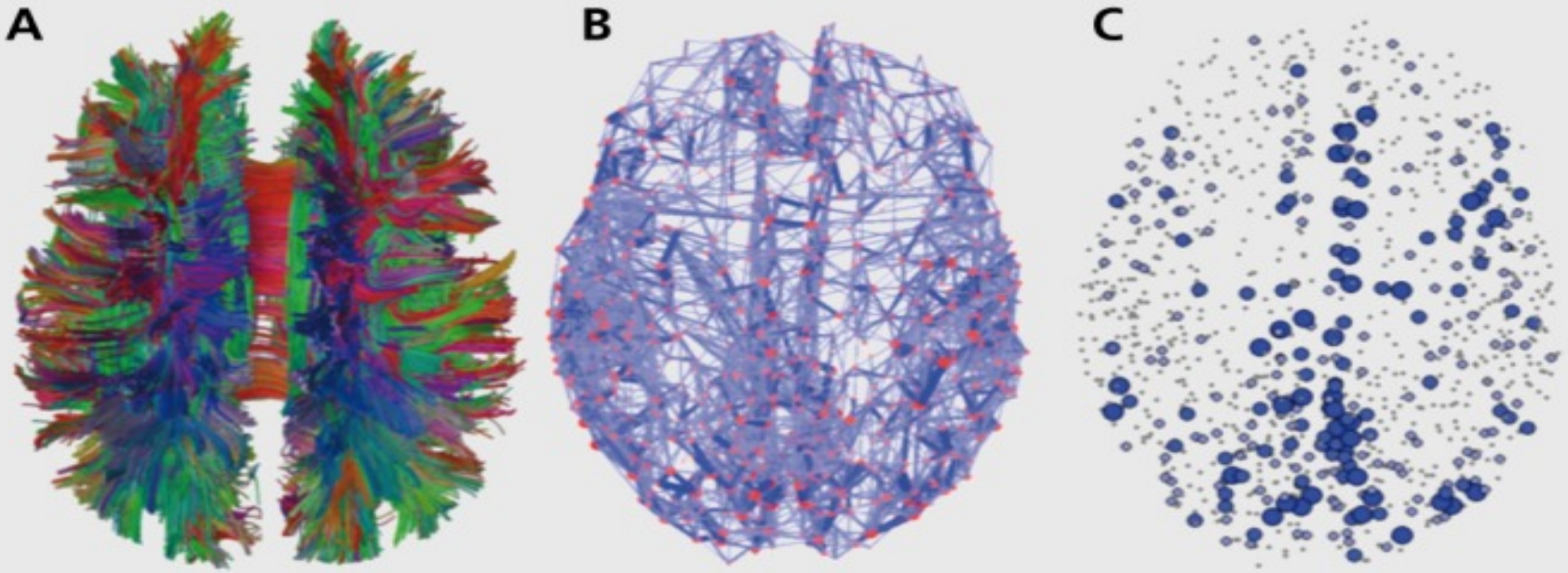
Class schedule & overview of the class

https://paballand.github.io/teaching/on.html

What is network thinking?

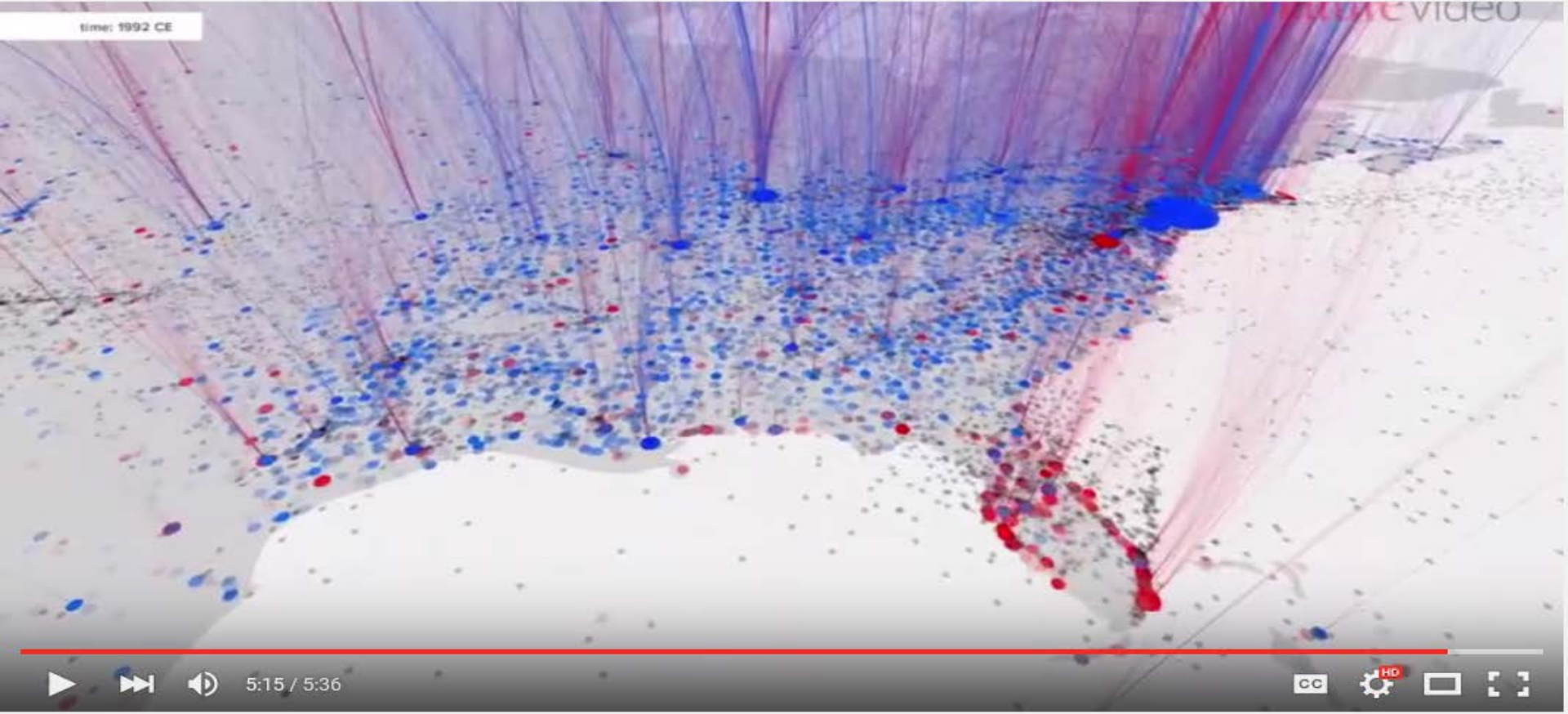
- A network-based paradigm is taking science by storm (Barabási, 2012)

Network structure of the brain



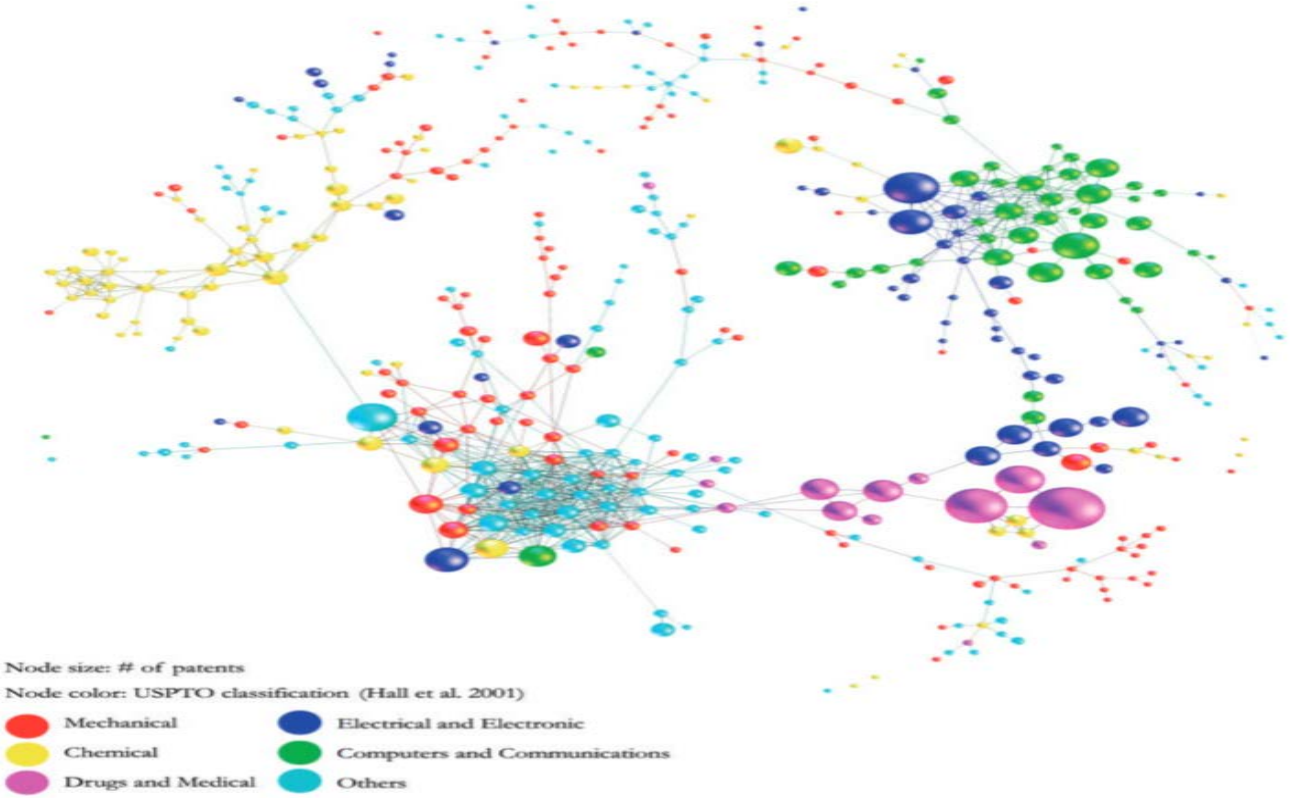
Sporns (2013)

Migration flows



Knowledge flows

Knowledge relatedness



Boschma, Balland and Kogler (2013)

What is network thinking?

- A network-based paradigm is taking science by storm (Barabási, 2012)...but also business

An interesting patent

US006285999B1

(12) **United States Patent**
Page

(10) **Patent No.: US 6,285,999 B1**
(45) **Date of Patent: Sep. 4, 2001**

(54) **METHOD FOR NODE RANKING IN A LINKED DATABASE**

(75) Inventor: **Lawrence Page**, Stanford, CA (US)

(73) Assignee: **The Board of Trustees of the Leland Stanford Junior University**, Stanford, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **09/004,827**

(22) Filed: **Jan. 9, 1998**

Related U.S. Application Data

(60) Provisional application No. 60/035,205, filed on Jan. 10, 1997.

(51) **Int. Cl.**[7] **G06F 17/30**
(52) **U.S. Cl.** **707/5**; 707/7; 707/501
(58) **Field of Search** 707/100, 5, 7, 707/513, 1–3, 10, 104, 501; 345/440; 382/226, 229, 230, 231

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,953,106	*	8/1990	Gansner et al.	345/440
5,450,535	*	9/1995	North	395/140
5,748,954		5/1998	Mauldin	395/610
5,752,241	*	5/1998	Cohen	707/3
5,832,494	*	11/1998	Egger et al.	707/102
5,848,407	*	12/1998	Ishikawa et al.	707/2
6,014,678	*	1/2000	Inoue et al.	707/501

OTHER PUBLICATIONS

S. Jeromy Carriere et al, "Web Query: Searching and Visualizing the Web through Connectivity", Computer Networks and ISDN Systems 29 (1997), pp. 1257–1267.*

Wang et al "Prefetching in Worl Wide Web", IEEE 1996, pp. 28–32.*

Ramer et al "Similarity, Probability and Database Organisation: Extended Abstract", 1996, pp. 272.276.*

Craig Boyle "To link or not to link: An empirical comparison of Hypertext linking strategies". ACM 1992, pp. 221–231.*

L. Katz, "A new status index derived from sociometric analysis," 1953, Psychometricka, vol. 18, pp. 39–43.

C.H. Hubbell, "An input–output approach to clique identification sociometry," 1965, pp. 377–399.

Mizruchi et al., "Techniques for disaggregating centrality scores in social networks," 1996, Sociological Methodology, pp. 26–48.

E. Garfield, "Citation analysis as a tool in journal evaluation," 1972, Science, vol. 178, pp. 471–479.

Pinski et al., "Citation influence for journal aggregates of scientific publications: Theory, with application to the literature of physics," 1976, Inf. Proc. And Management, vol. 12, pp. 297–312.

N. Geller, "On the citation influence methodology of Pinski and Narin," 1978, Inf. Proc. And Management, vol. 14, pp. 93–95.

P. Doreian, "Measuring the relative standing of disciplinary journals," 1988, Inf. Proc. And Management, vol. 24, pp. 45–56.

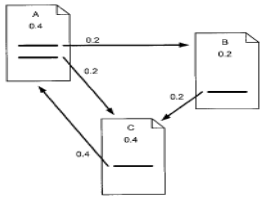
(List continued on next page.)

Primary Examiner—Thomas Black
Assistant Examiner—Uyen Le
(74) *Attorney, Agent, or Firm*—Harrity & Snyder L.L.P.

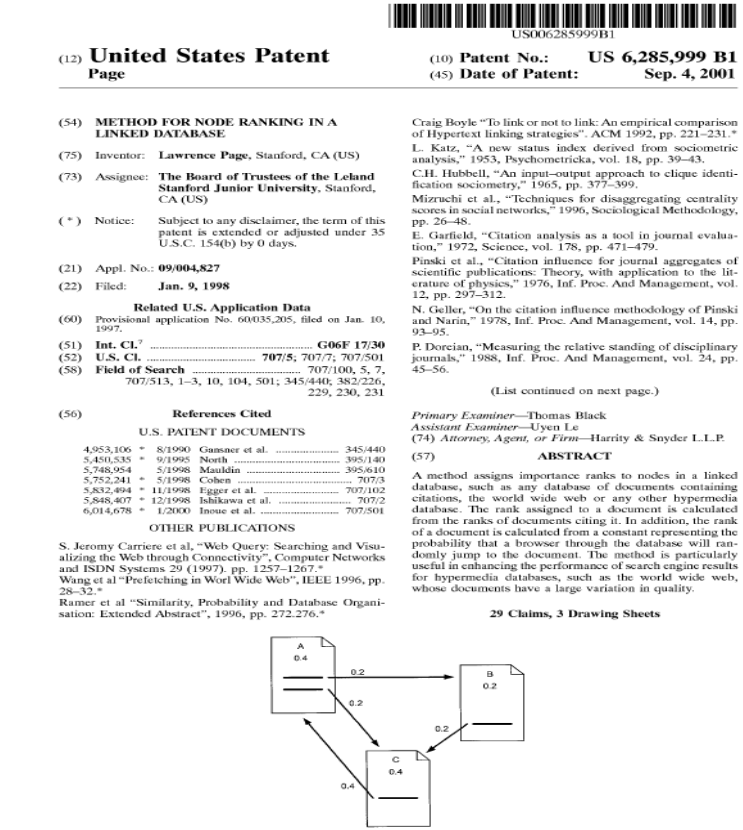
(57) **ABSTRACT**

A method assigns importance ranks to nodes in a linked database, such as any database of documents containing citations, the world wide web or any other hypermedia database. The rank assigned to a document is calculated from the ranks of documents citing it. In addition, the rank of a document is calculated from a constant representing the probability that a browser through the database will randomly jump to the document. The method is particularly useful in enhancing the performance of search engine results for hypermedia databases, such as the world wide web, whose documents have a large variation in quality.

29 Claims, 3 Drawing Sheets



The Google PageRank algorithm



US006285999B1

(12) **United States Patent**
Page

(10) **Patent No.: US 6,285,999 B1**
(45) **Date of Patent: Sep. 4, 2001**

(54) **METHOD FOR NODE RANKING IN A LINKED DATABASE**

(75) Inventor: **Lawrence Page**, Stanford, CA (US)

(73) Assignee: **The Board of Trustees of the Leland Stanford Junior University**, Stanford, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **09/004,827**

(22) Filed: **Jan. 9, 1998**

Related U.S. Application Data

(60) Provisional application No. 60/035,205, filed on Jan. 10, 1997.

(51) **Int. Cl.**[7] **G06F 17/30**
(52) **U.S. Cl.** **707/5**; 707/7; 707/501
(58) **Field of Search** 707/100, 5, 7, 707/513, 1–3, 10, 104, 501; 345/440; 382/226, 229, 230, 231

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,953,106	*	8/1990	Gansner et al.	345/440
5,450,535	*	9/1995	North	395/140
5,748,954		5/1998	Mauldin	395/610
5,752,241	*	5/1998	Cohen	707/3
5,832,494	*	11/1998	Egger et al.	707/102
5,848,407	*	12/1998	Ishikawa et al.	707/2
6,014,678	*	1/2000	Inoue et al.	707/501

OTHER PUBLICATIONS

S. Jeromy Carriere et al, "Web Query: Searching and Visualizing the Web through Connectivity", Computer Networks and ISDN Systems 29 (1997), pp. 1257–1267.*
Wang et al "Prefeteching in Worl Wide Web", IEEE 1996, pp. 28–32.*
Ramer et al "Similarity, Probability and Database Organisation: Extended Abstract", 1996, pp. 272.276.*
Craig Boyle "To link or not to link: An empirical comparison of Hypertext linking strategies". ACM 1992, pp. 221–231.*
L. Katz, "A new status index derived from sociometric analysis," 1953, Psychometricka, vol. 18, pp. 39–43.
C.H. Hubbell, "An input–output approach to clique identification sociometry," 1965, pp. 377–399.
Mizruchi et al., "Techniques for disaggregating centrality scores in social networks," 1996, Sociological Methodology, pp. 26–48.
E. Garfield, "Citation analysis as a tool in journal evaluation," 1972, Science, vol. 178, pp. 471–479.
Pinski et al., "Citation influence for journal aggregates of scientific publications: Theory, with application to the literature of physics," 1976, Inf. Proc. And Management, vol. 12, pp. 297–312.
N. Geller, "On the citation influence methodology of Pinski and Narin," 1978, Inf. Proc. And Management, vol. 14, pp. 93–95.
P. Doreian, "Measuring the relative standing of disciplinary journals," 1988, Inf. Proc. And Management, vol. 24, pp. 45–56.

(List continued on next page.)

Primary Examiner—Thomas Black
Assistant Examiner—Uyen Le
(74) *Attorney, Agent, or Firm*—Harrity & Snyder L.L.P.

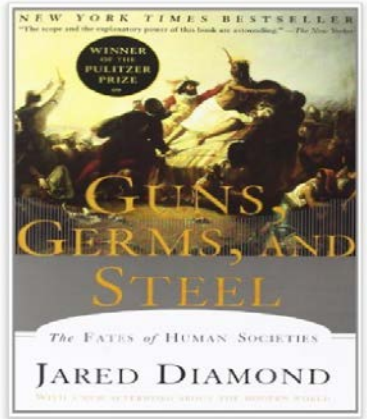
(57) **ABSTRACT**

A method assigns importance ranks to nodes in a linked database, such as any database of documents containing citations, the world wide web or any other hypermedia database. The rank assigned to a document is calculated from the ranks of documents citing it. In addition, the rank of a document is calculated from a constant representing the probability that a browser through the database will randomly jump to the document. The method is particularly useful in enhancing the performance of search engine results for hypermedia databases, such as the world wide web, whose documents have a large variation in quality.

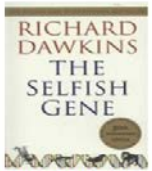
29 Claims, 3 Drawing Sheets

Google

Amazon

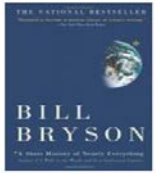


Customers Who Bought This Item Also Bought

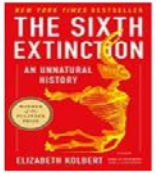


The Selfish Gene: 30th Anniversary Edition--with a new Introduction by the...
› Richard Dawkins
★★★★½ 707
#1 Best Seller in Genetics
Paperback

1493: Uncovering the New World Columbus Created
› Charles C. Mann
★★★★½ 513
Paperback
$10.08 Prime



A Short History of Nearly Everything
› Bill Bryson
★★★★½ 2,179
Paperback
$9.60 Prime



The Sixth Extinction: An Unnatural History
› Elizabeth Kolbert
★★★★½ 1,174
#1 Best Seller in Natural History
Paperback

Facebook recommendation

People You May Know Add Friend Remove

Social networks and population mapping

Paul Butler (Facebook - 2010)

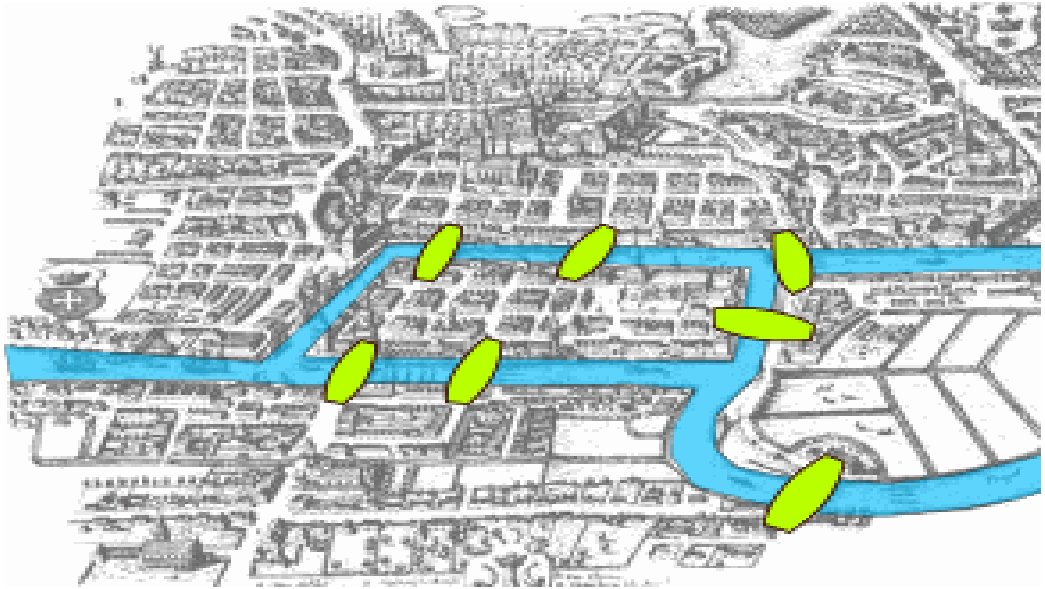
What is network thinking?

- A network-based paradigm is taking science by storm (Barabási, 2012)…but also business

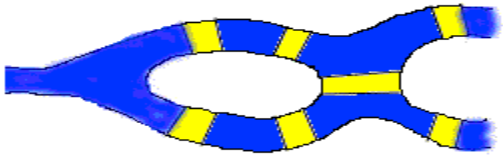
- Network analysis is a broad intellectual approach instead of a narrow set of methods (Wellman, 1983)

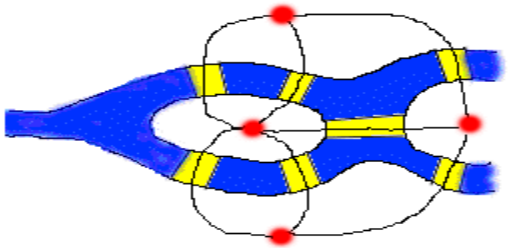
The seven bridges of Königsberg

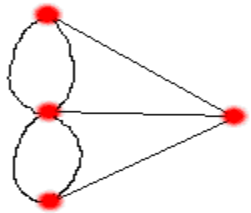
Fundamental problem in the history of mathematics : find a walk through the city that would cross each bridge once and only once



Leonhard Euler solution (1735)







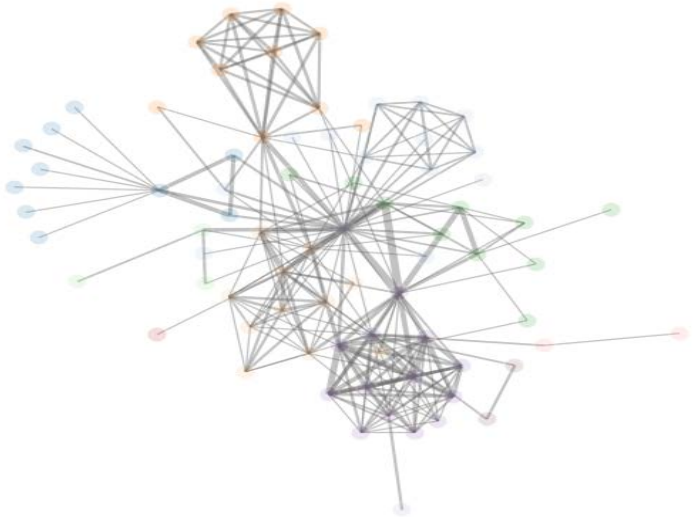
- The route inside each land mass is irrelevant, only the patterns of connection are important
- Abstract reformulation: collapse areas of land separated by the river into points (nodes) connected by the 7 bridges (edges)
- Euler used a network (graph-based) approach to prove that there is no path that would cross each bridge once and only once
- Foundation of graph theory and mathematical topology

What is network thinking?

- A network-based paradigm is taking science by storm (Barabási, 2012)…but also business
- Network analysis is a broad intellectual approach instead of a narrow set of methods (Wellman, 1983)
- A network-based paradigm shifts the unit of analysis from **individuals** and their attributes to (the structure of) their **relationships**

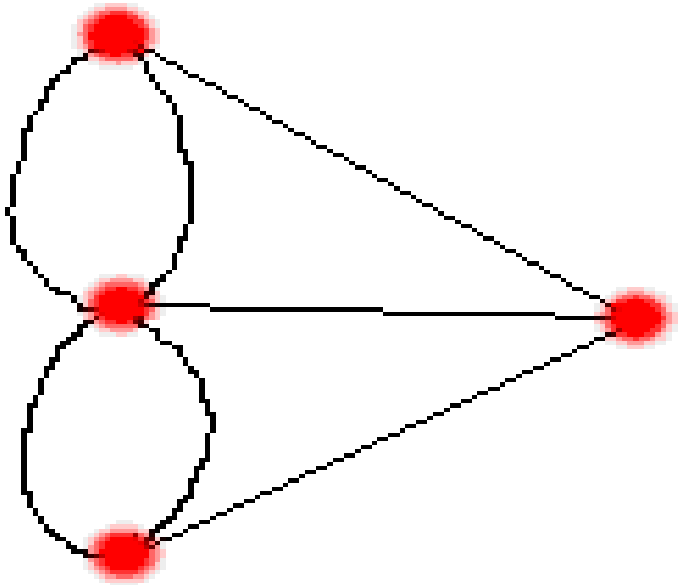
Network metrics & visualization

- Network centrality
- Brokerage
- Network density
- Core-periphery structure
- Average path length
- Clustering coefficient
- Communities
- Degree distribution
- Statistical model of network dyna
- ...

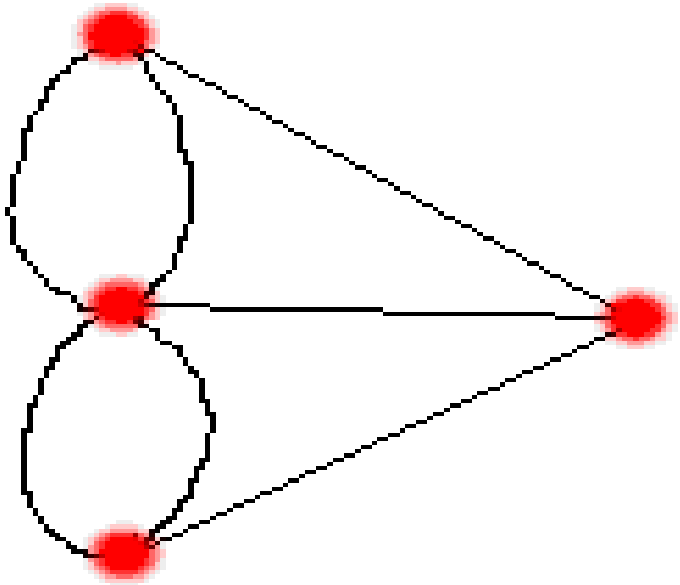


Network thinking in music

Network thinking reduces complexity



Hidden properties of network structures



Thanks!

paballand.com

github.com/PABalland/EconGeo

Computer lab: R & RStudio

- In this course we will perform structural network analysis with packages implemented in the R statistical software
- R is the software – but we will use Rstudio as an interface
- R is an open-source project lifted by a virtual community of thousands of developers and million of users worldwide

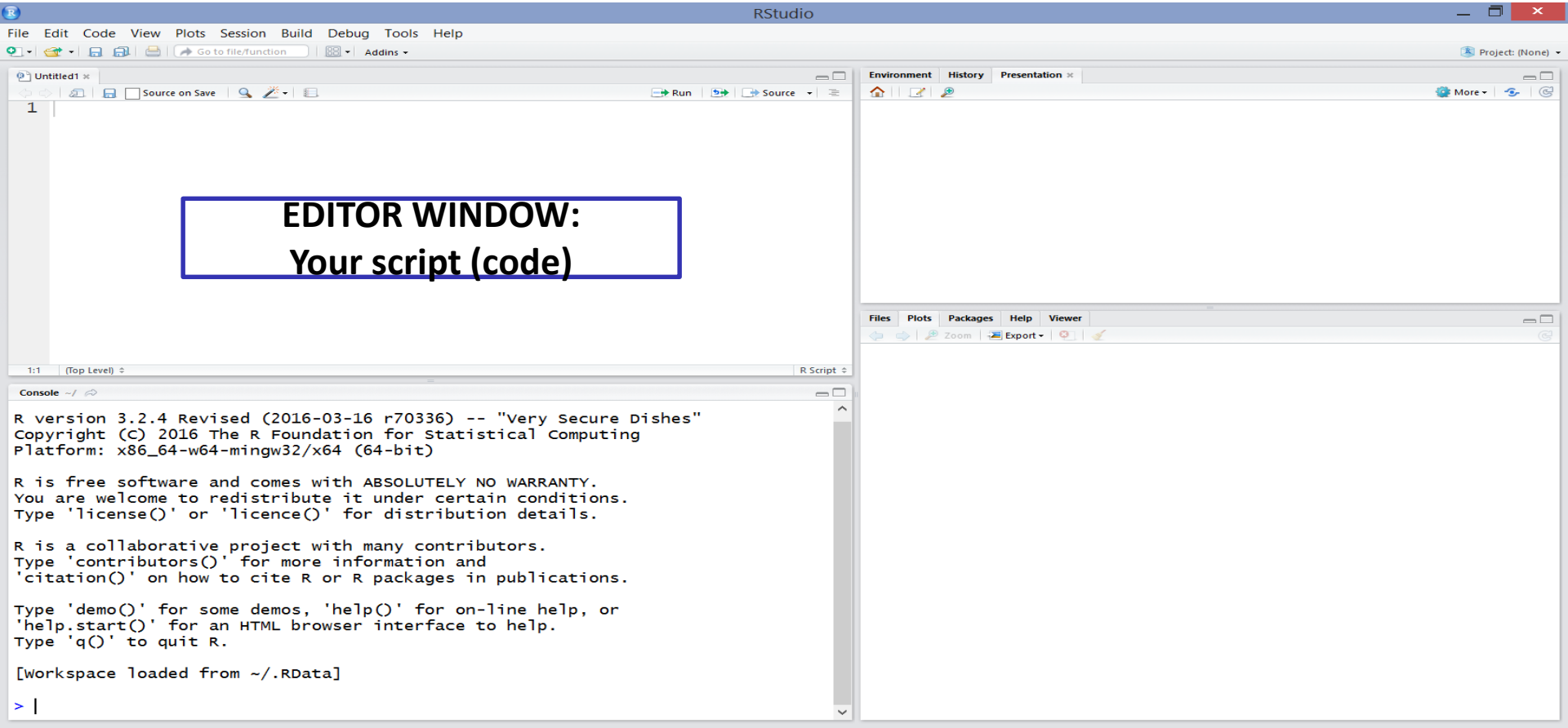
Why R?

- Reproducibility – R scripts
- Today R offers the most elegant and comprehensive language for the structural and dynamic analysis of networks
- It's free and contains state-of-the-art statistical and graphical routines not yet available in other software
- You can do all your analysis in R, but also data scrapping, create a webpage, or write your research paper

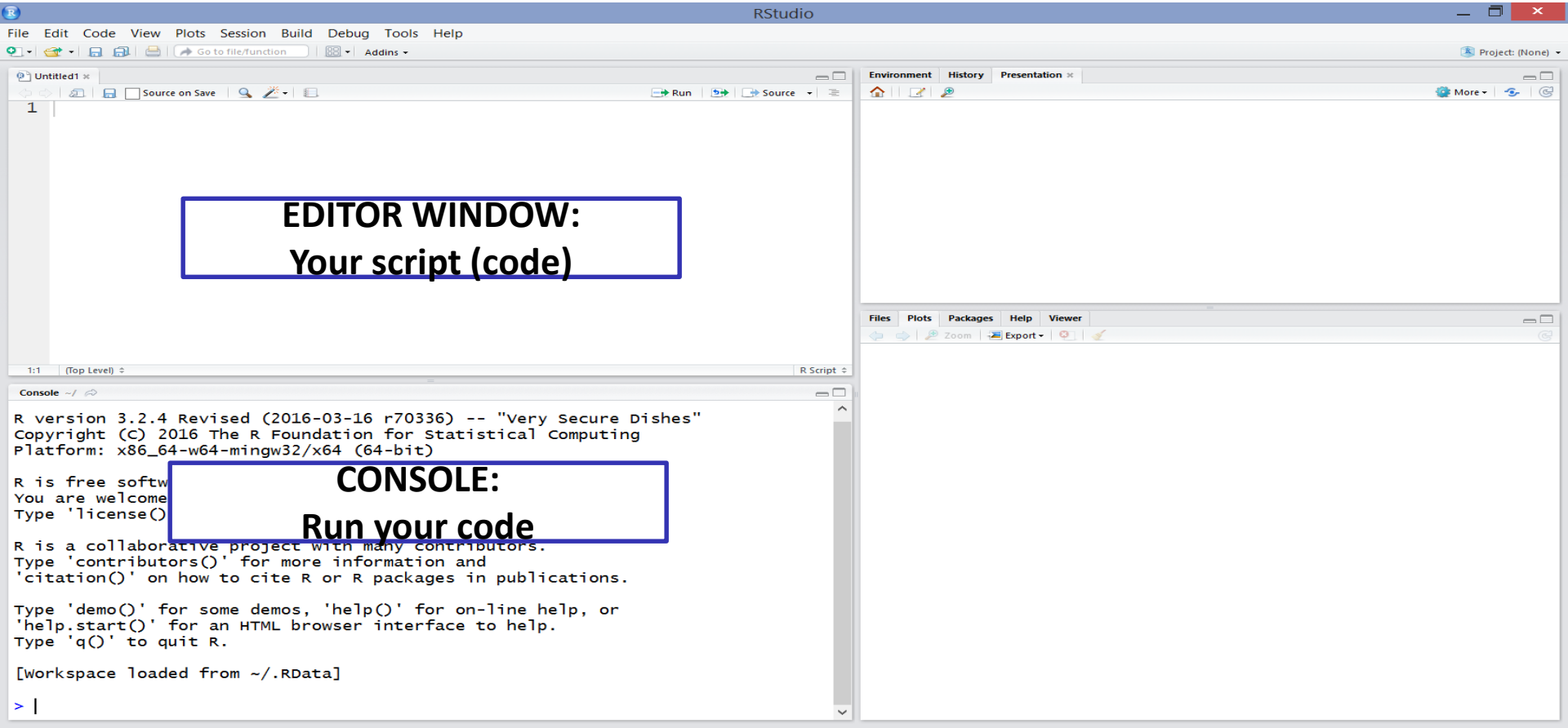
Getting started with R

- Using R is easier than it looks like. And once you master it, you save a ridiculous amount of time
- Afraid of R? It is just a big calculator (a very smart one)
- R is case sensitive
- The # character at the beginning of a line signifies a comment, it is ignored by R
- The <- operator assigns values to objects (data structures)

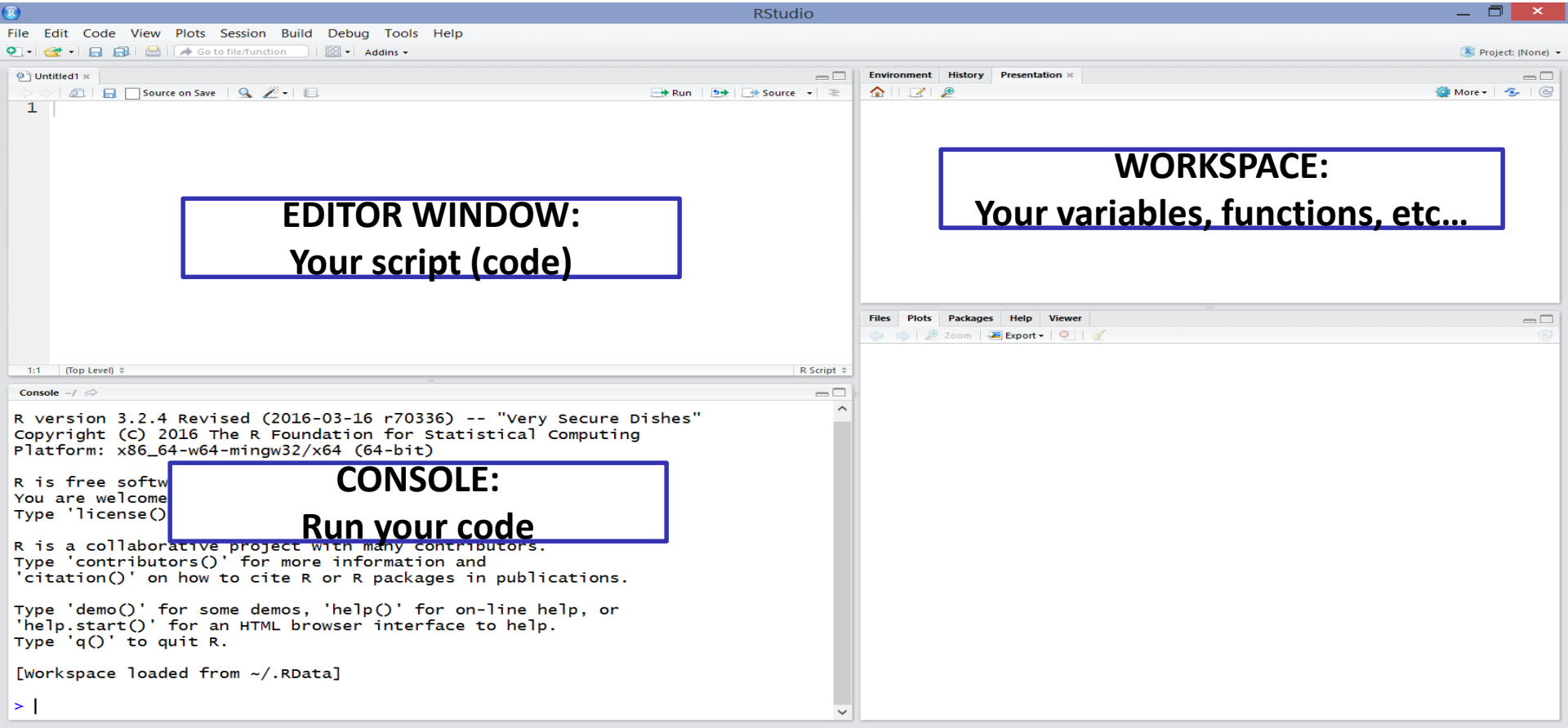
RStudio


RStudio
File Edit Code View Plots Session Build Debug Tools Help
Untitled1
Source on Save
Run
Source
1
EDITOR WINDOW:
Your script (code)
1:1 (Top Level)
R Script
Environment History Presentation
More
Files Plots Packages Help Viewer
Zoom
Export
Console ~/
R version 3.2.4 Revised (2016-03-16 r70336) -- "Very Secure Dishes"
Copyright (C) 2016 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
[Workspace loaded from ~/.RData]
>

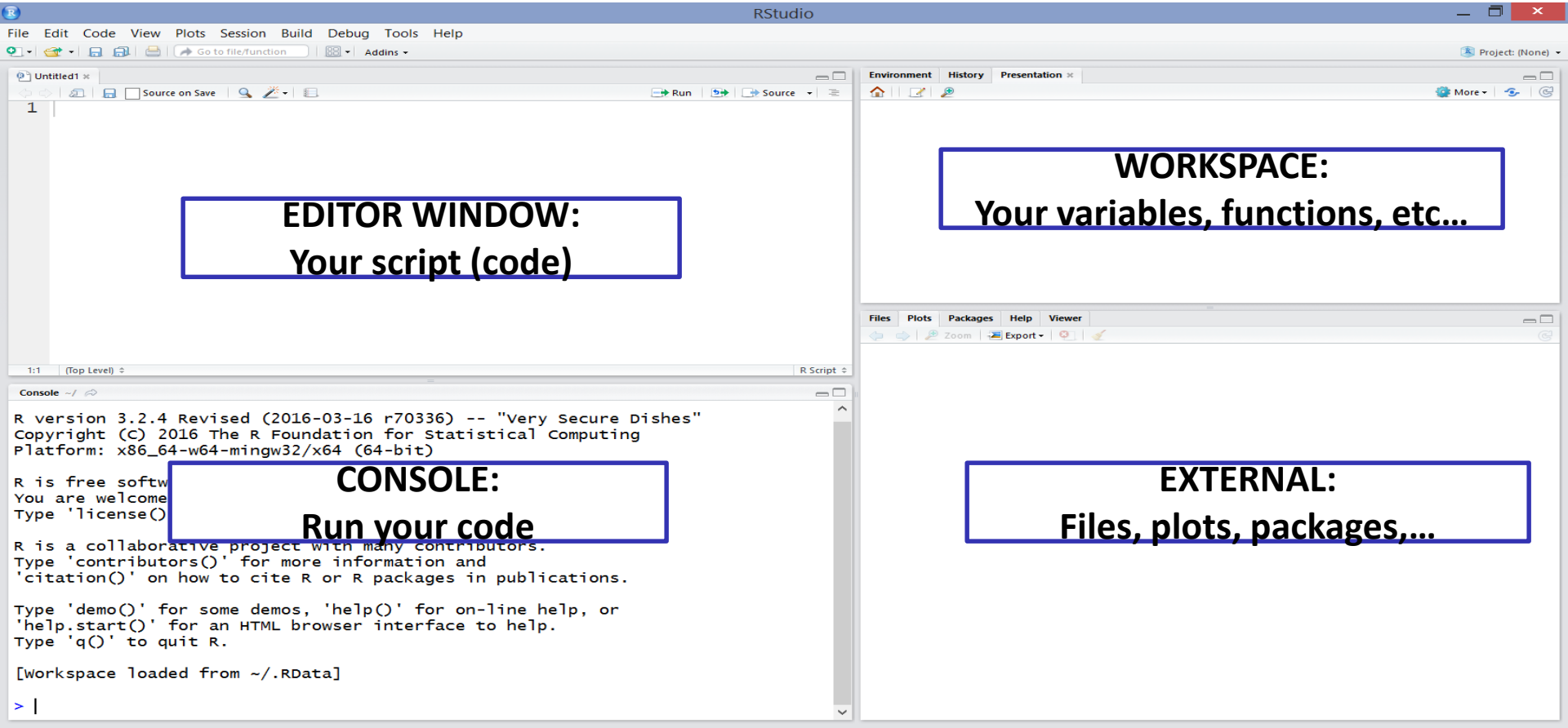
RStudio


RStudio
File Edit Code View Plots Session Build Debug Tools Help
Untitled1
Source on Save
Run
Source
1
EDITOR WINDOW:
Your script (code)
Environment History Presentation
More
Files Plots Packages Help Viewer
Zoom
Export
Console ~/
R version 3.2.4 Revised (2016-03-16 r70336) -- "Very Secure Dishes"
Copyright (C) 2016 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)
R is free softw
You are welcome
Type 'license()
CONSOLE:
Run your code
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
[Workspace loaded from ~/.RData]
>

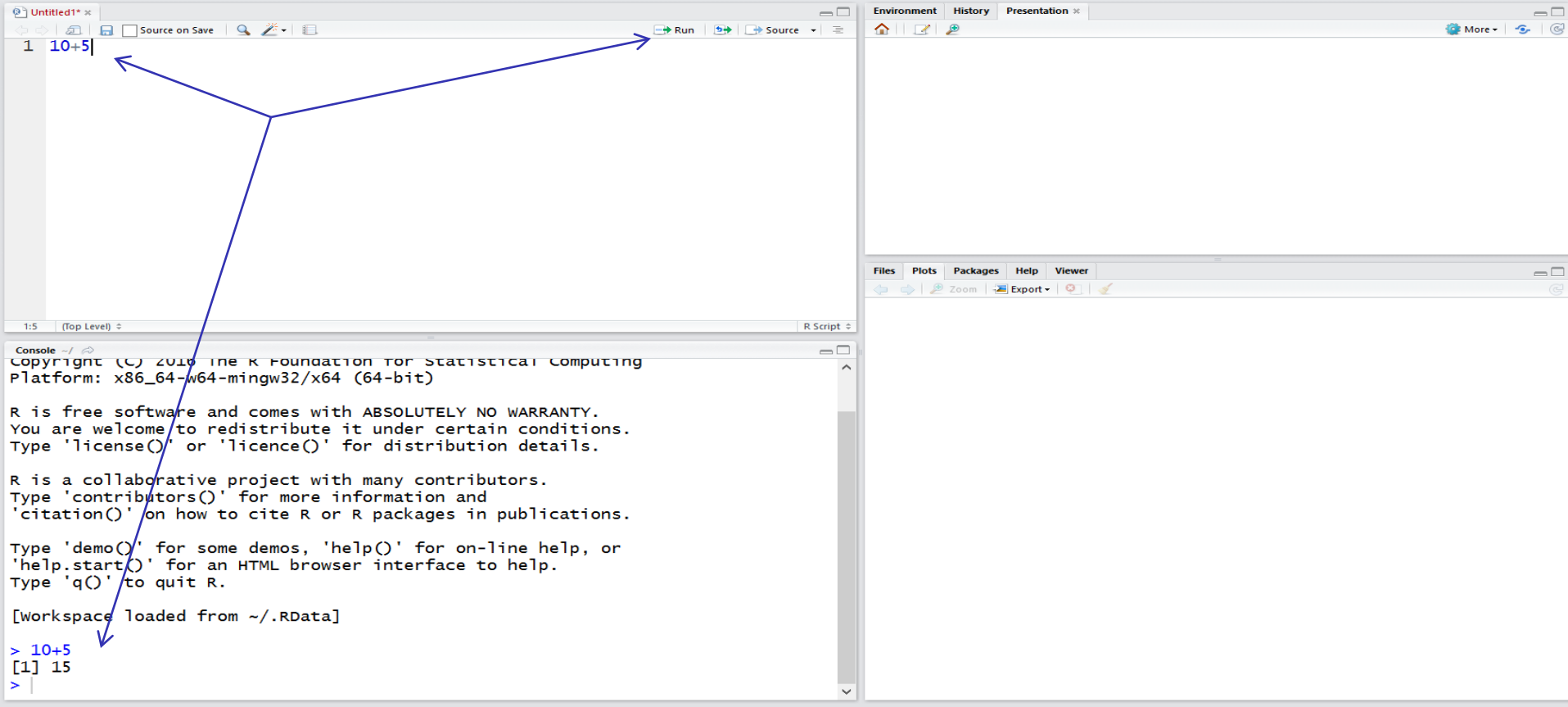
RStudio


RStudio
File Edit Code View Plots Session Build Debug Tools Help
Untitled1
Source on Save
Run
Source
1
EDITOR WINDOW:
Your script (code)
Environment History Presentation
More
WORKSPACE:
Your variables, functions, etc…
Files Plots Packages Help Viewer
Zoom
Export
Console ~/
R version 3.2.4 Revised (2016-03-16 r70336) -- "Very Secure Dishes"
Copyright (C) 2016 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)
R is free softw
You are welcome
Type 'license()
CONSOLE:
Run your code
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
[Workspace loaded from ~/.RData]
>

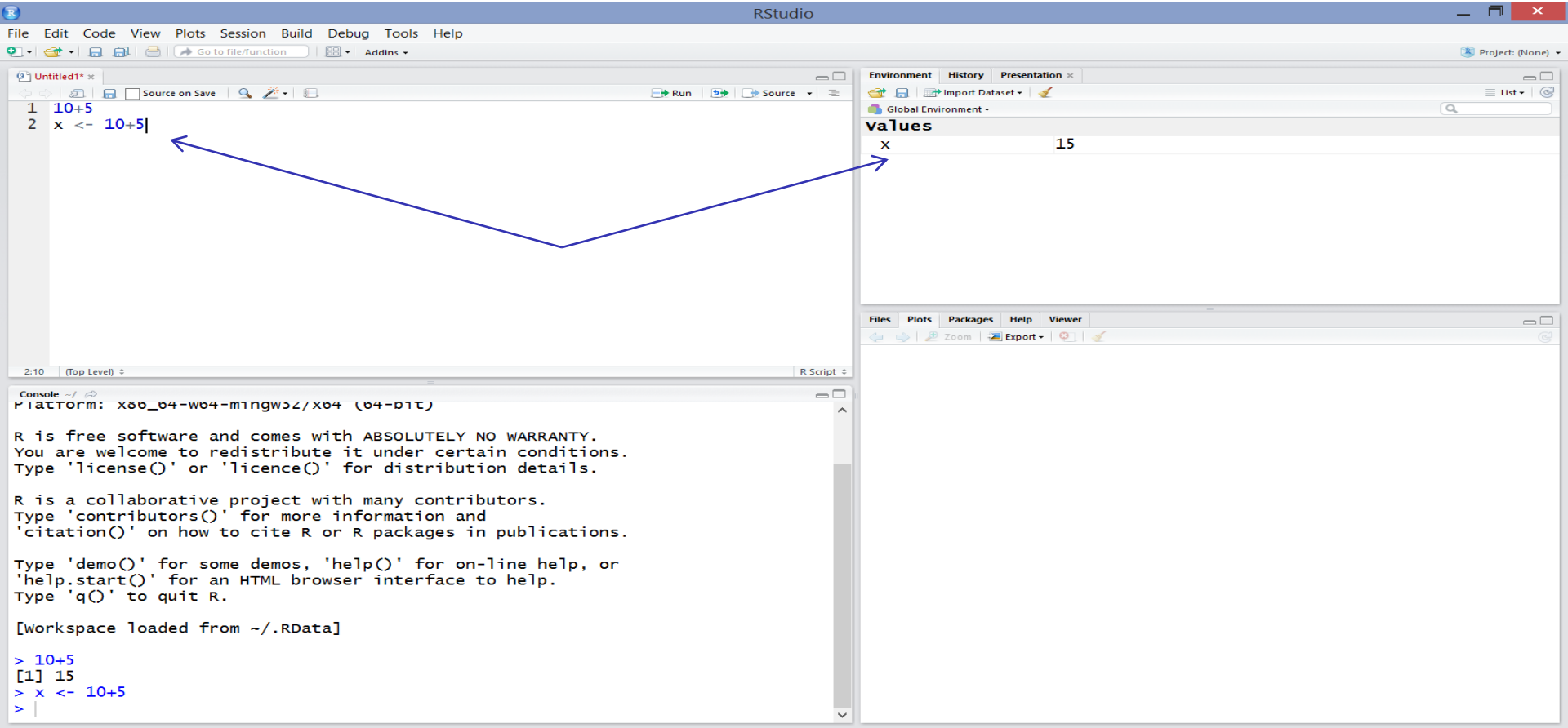
RStudio


RStudio
File Edit Code View Plots Session Build Debug Tools Help
Untitled1
Source on Save
Run
Source
1
EDITOR WINDOW:
Your script (code)
Environment History Presentation
More
WORKSPACE:
Your variables, functions, etc…
Files Plots Packages Help Viewer
Zoom
Export
EXTERNAL:
Files, plots, packages,…
Console ~/
R version 3.2.4 Revised (2016-03-16 r70336) -- "Very Secure Dishes"
Copyright (C) 2016 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)
R is free softw
You are welcome
Type 'license()
CONSOLE:
Run your code
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
[Workspace loaded from ~/.RData]
>

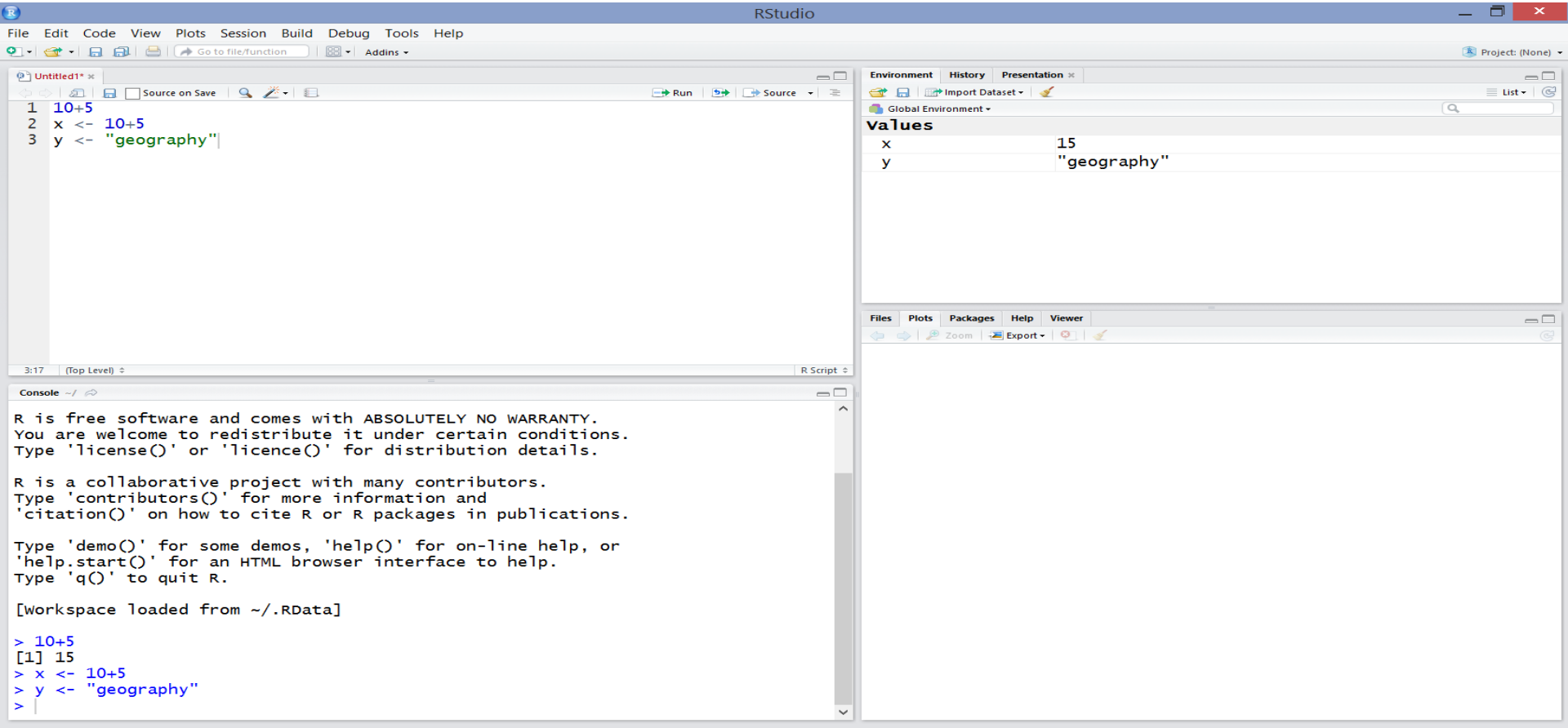
Let’s get started



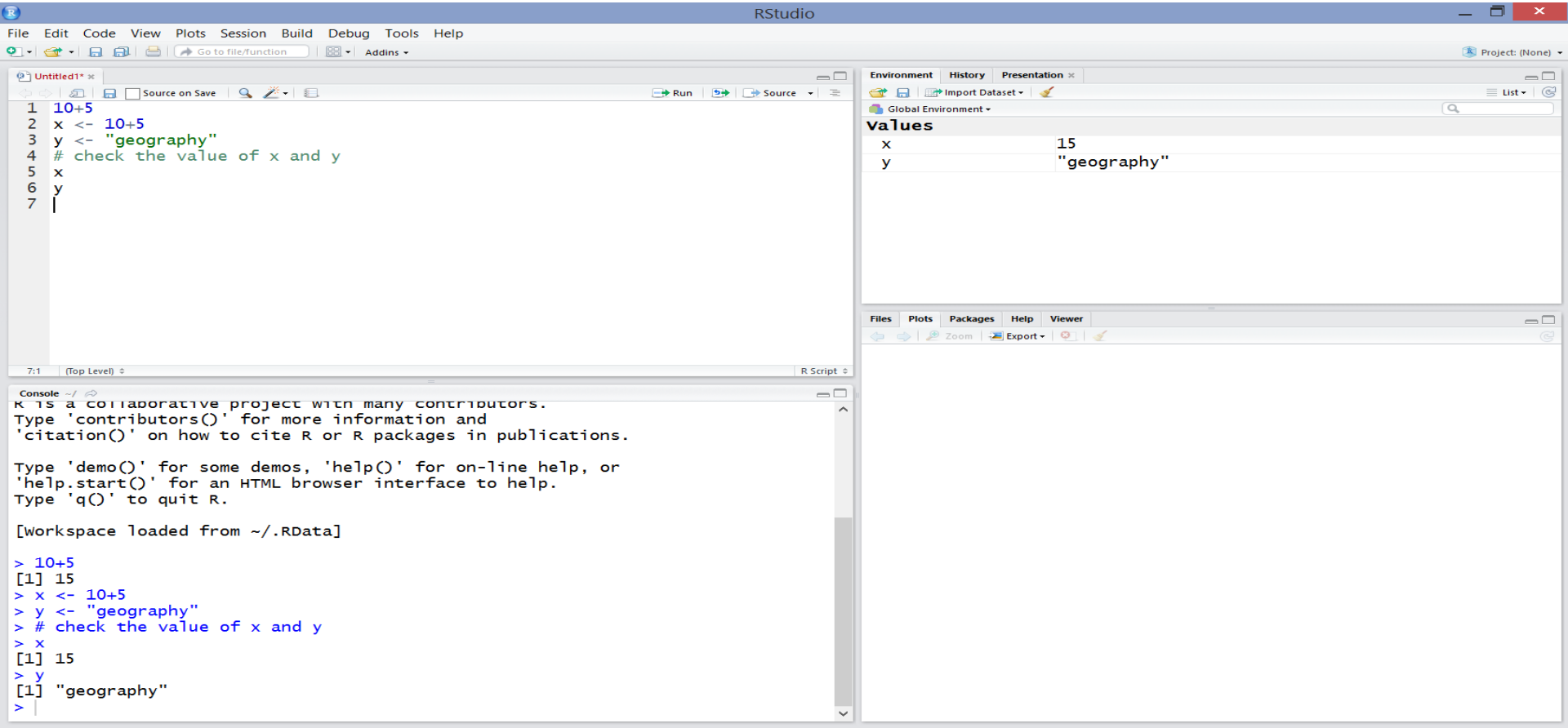
Create a variable “x”



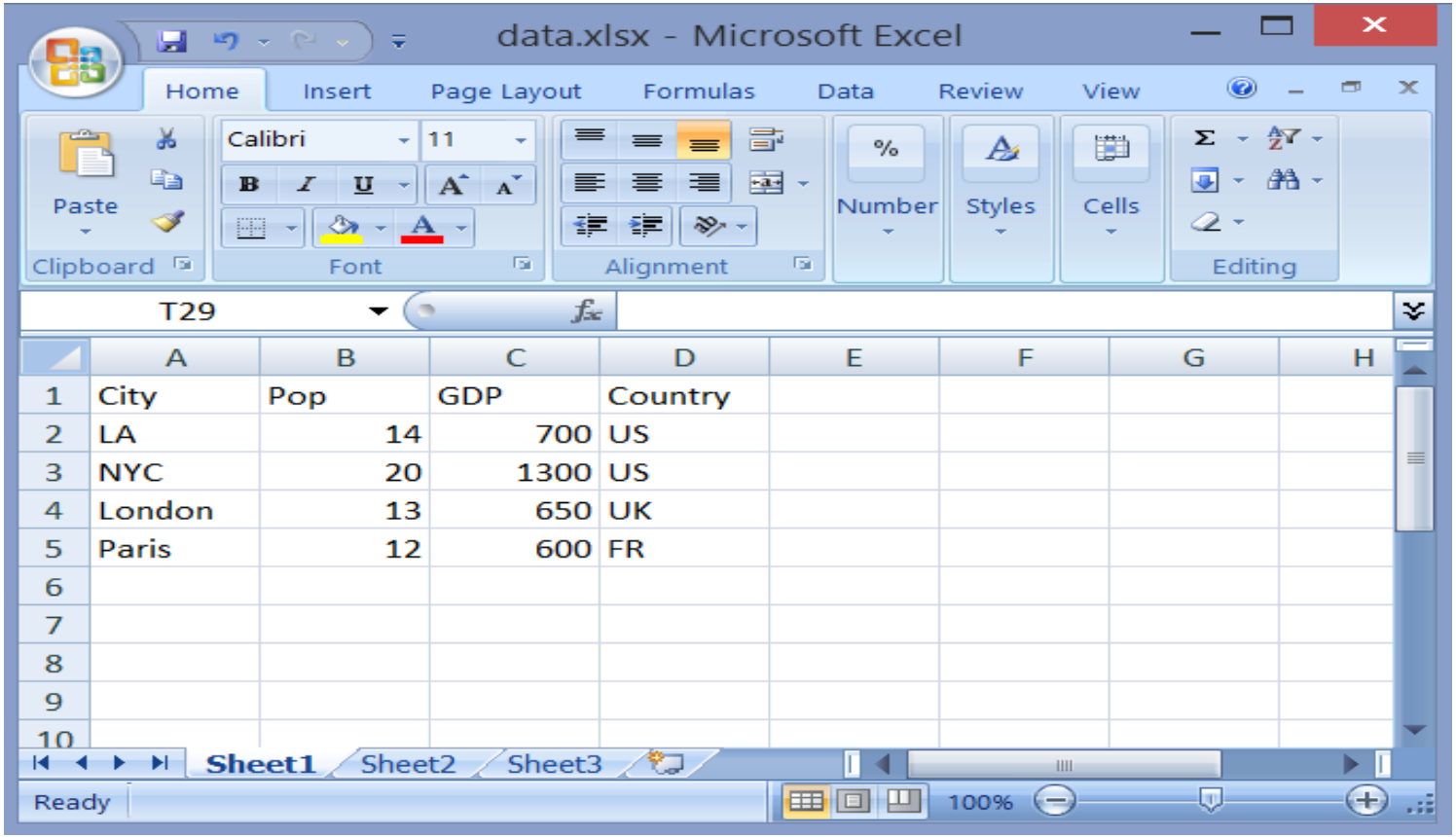
Create a variable y



Check the value of x and y

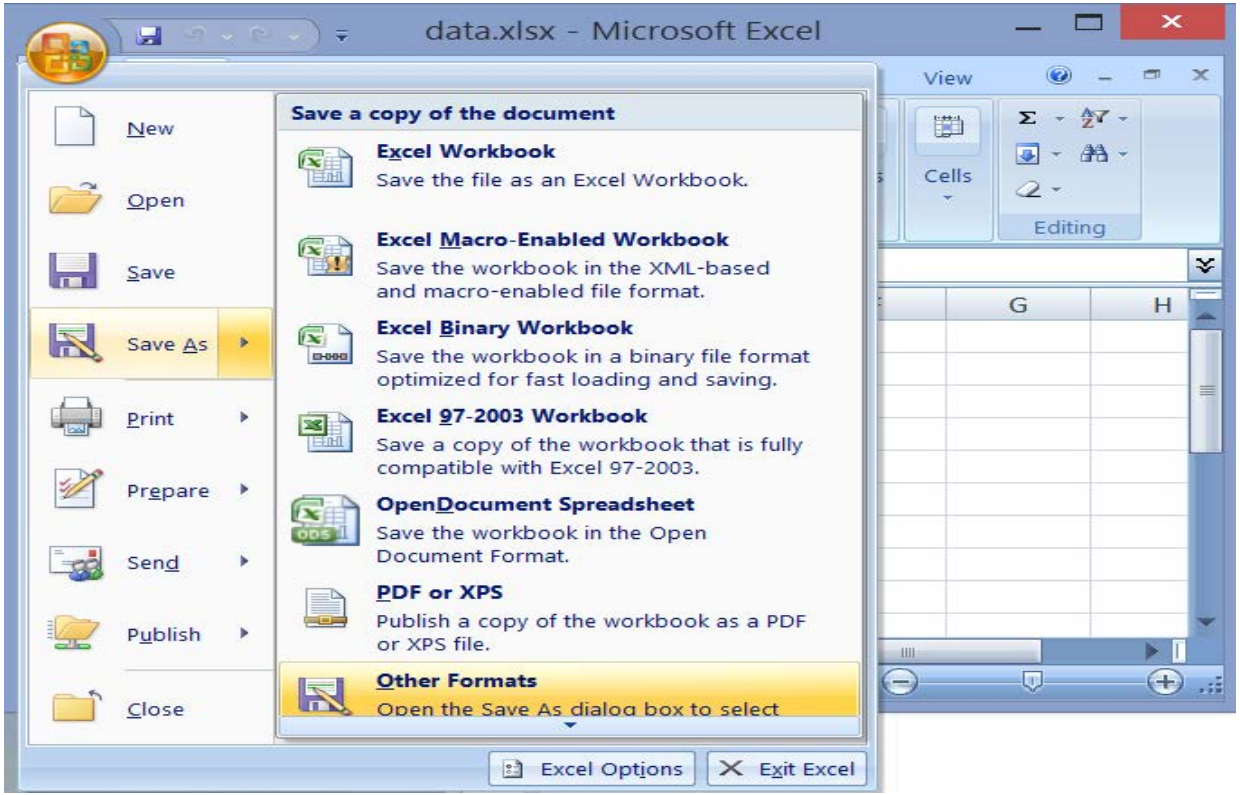


Let’s create a toy dataset

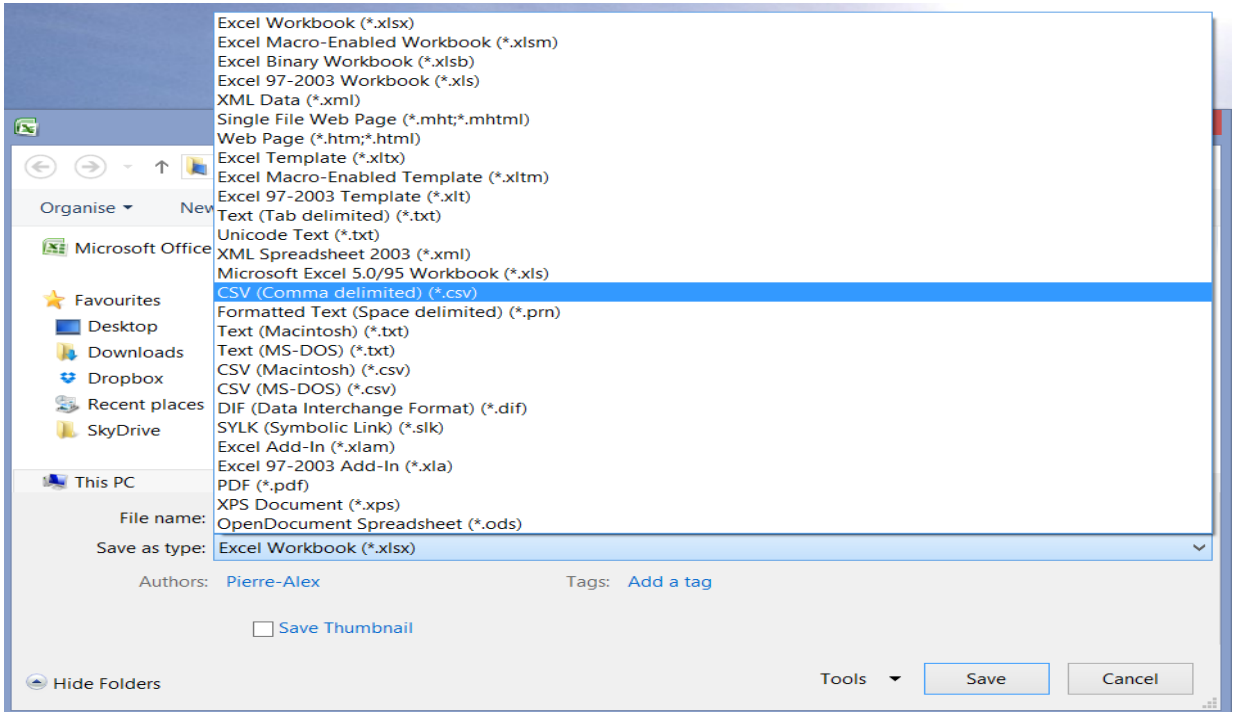


	A	B	C	D
1	City	Pop	GDP	Country
2	LA	14	700	US
3	NYC	20	1300	US
4	London	13	650	UK
5	Paris	12	600	FR

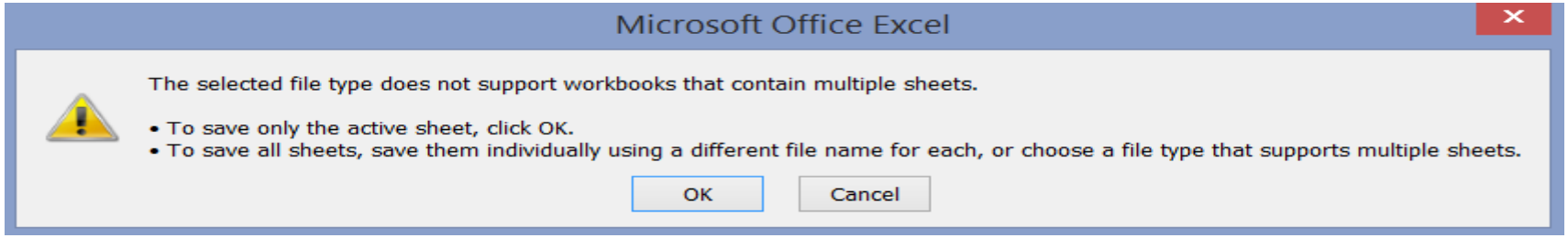
Save as a .csv file



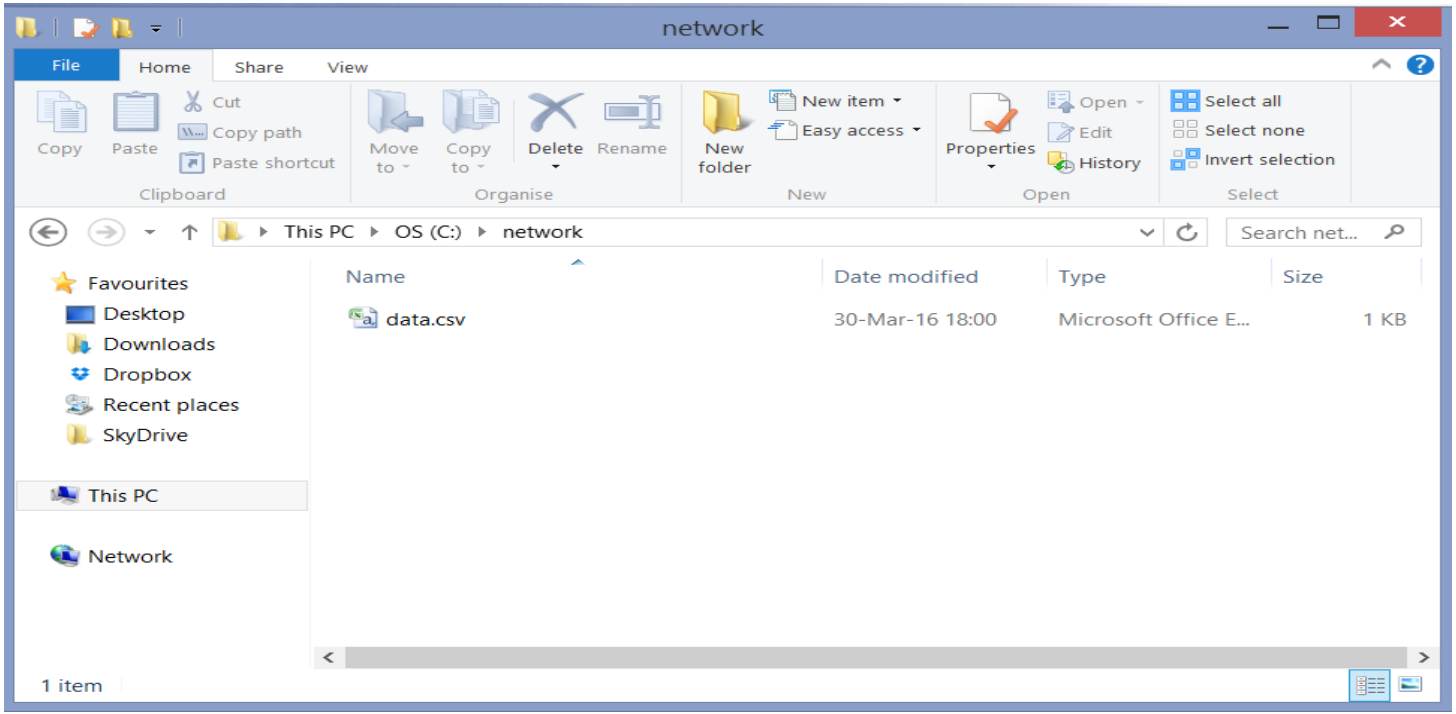
Save as a .csv file



Two warnings: ok



Create a new folder and move the .csv



This is your file path