

Crush course

# Introduction to practical biostatistical analysis with R

Sergey Mastitsky ©

Klaipeda, 28-30 September 2011

# Five statistical things every biologist should know

*Ewan Birney, Head of Nucleotide Data at European Bioinformatics Institute, EMBL*

From a [post](#) in his blog “Five statistical things I wished I had been taught 20 years ago”

- Nonparametric statistics
- R
- The problem of multiple testing
- The relationship between P-value, effect size, and sample size
- Linear models and PCA

# Why R?

- Absolutely free – very important for Academia
- Designed for statistical analysis
- Virtually unlimited options for data visualization
- Over 3000 add-on packages, for all imaginable sorts of analysis
- Develops faster than any commercial software
- Hundreds of books and other literature published
- Huge community – estimated 2 million users (=> support and bug fixing)

# R drawbacks

- Works from the command line – not friendly for the first-time users
- R is a language – takes time to learn
- Help documentation is often very technical – not easy to understand even for experienced users

# How to learn R?

- Books – most are published by *Springer*, *Cambridge University Press*, and *Chapman & Hall*
- Blogs – extremely useful (and free!)
- Special courses – like this one 😊; also consider [www.statistics.com](http://www.statistics.com)

# Major topics of this (practice-oriented) course:

- Main principles of work with R and RStudio
- Descriptive statistics, basics of R graphics, classical statistical tests (e.g., t-test, ANOVA, correlation analysis)
- Linear modeling with R (multiple regression and logistic regression)
- In the end: you will have all the course materials, including re-usable code examples

# There will be a combination of:

- I explain the theory and show real-time examples – you repeat
- I give you assignments – you work on them
- You work on individual projects in pairs
- In the end: you present your individual projects and we discuss them

# Introduction to practical biostatistical analysis with R

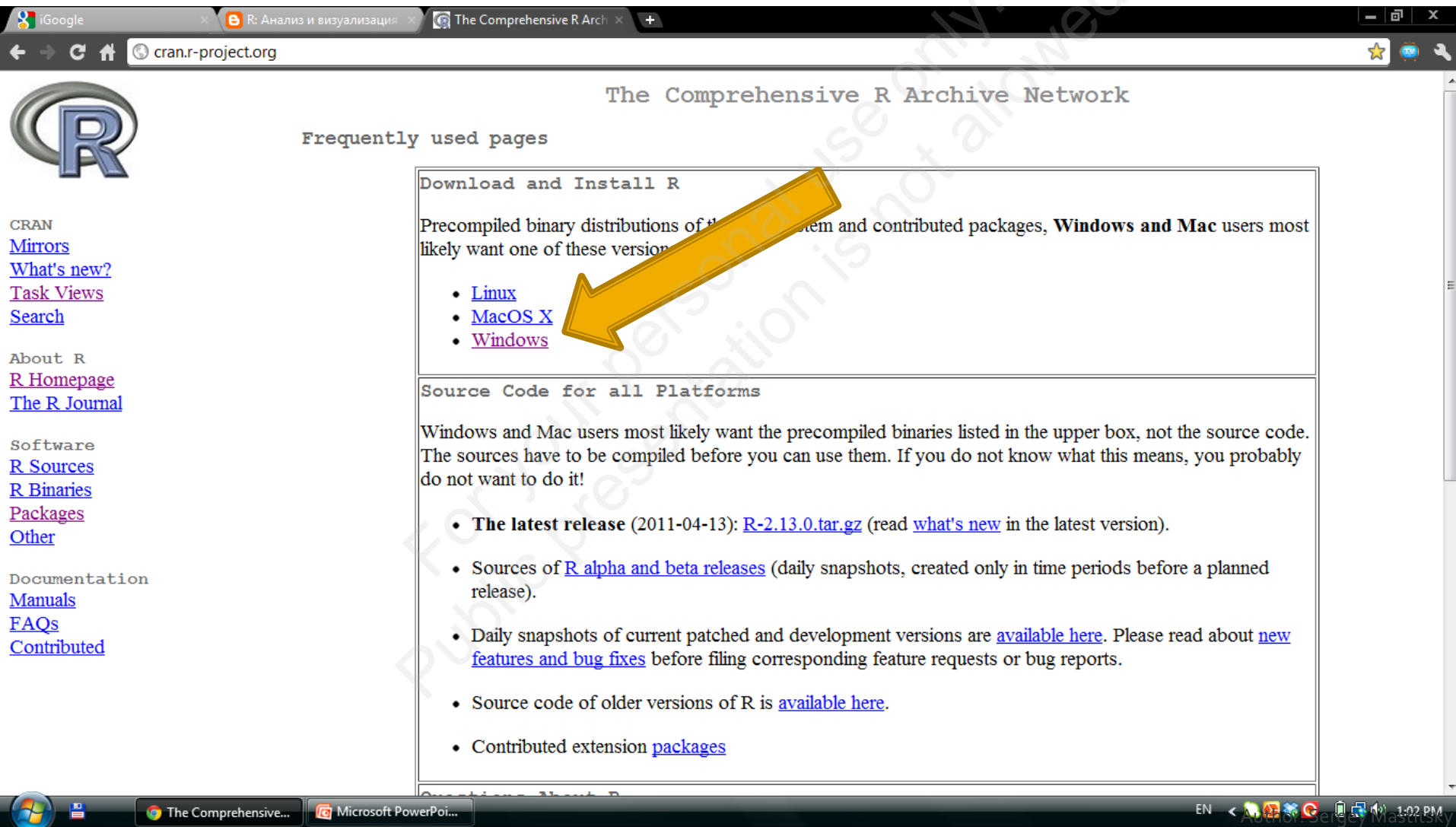
Topic 1

## Installation of R and RStudio

For your personal use only.  
Public presentation is not allowed



# CRAN website: <http://cran.r-project.org>



The screenshot shows a web browser window with the address bar containing [cran.r-project.org](http://cran.r-project.org). The page title is "The Comprehensive R Archive Network". The main content area is titled "Frequently used pages" and contains two sections:

- Download and Install R**

Precompiled binary distributions of the system and contributed packages, **Windows and Mac** users most likely want one of these versions:

  - [Linux](#)
  - [MacOS X](#)
  - [Windows](#)
- Source Code for all Platforms**

Windows and Mac users most likely want the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

  - **The latest release** (2011-04-13): [R-2.13.0.tar.gz](#) (read [what's new](#) in the latest version).
  - Sources of [R alpha and beta releases](#) (daily snapshots, created only in time periods before a planned release).
  - Daily snapshots of current patched and development versions are [available here](#). Please read about [new features and bug fixes](#) before filing corresponding feature requests or bug reports.
  - Source code of older versions of R is [available here](#).
  - Contributed extension [packages](#)

A large yellow arrow points to the [Windows](#) link in the "Download and Install R" section.

# R installation package



## R for Windows

Subdirectories:

[base](#) ← the distribution (managed by Duncan Murdoch)  
[contrib](#) Binaries of contributed packages (managed by Uwe Ligges)

Please do not submit binaries to CRAN. Package developers might want to contact Duncan Murdoch or Uwe Ligges directly in case of questions / suggestions related to Windows binaries.

You may also want to read the [R FAQ](#) and [R for Windows FAQ](#).

Note: CRAN does some checks on these binaries for viruses, but cannot give guarantees. Use the normal precautions with downloaded executables.

CRAN

[Mirrors](#)  
[What's new?](#)  
[Task Views](#)  
[Search](#)

About R

[R Homepage](#)  
[The R Journal](#)

Software

[R Sources](#)  
[R Binaries](#)  
[Packages](#)  
[Other](#)

Documentation

[Manuals](#)  
[FAQs](#)  
[Contributed](#)

# R installation package



The screenshot shows a web browser window with the address bar displaying "cran.r-project.org". The page title is "R-2.13.0 for Windows (32/64 bit)". The main content area features a large yellow arrow pointing to the link "Download R 2.13.0 for Windows (37 megabytes, 32/64 bit)". Below this link are two sub-links: "Installation and other instructions" and "New features in this version: Windows specific, all platforms." The left sidebar contains navigation links for "CRAN", "About R", "Software", and "Documentation". The main content area also includes a section for "Frequently asked questions" with three bullet points, a note about the R FAQ, and a section for "Other builds" with three bullet points. At the bottom, there is a note to webmasters and a "Last change" timestamp.

R-2.13.0 for Windows (32/64 bit)

[Download R 2.13.0 for Windows](#) (37 megabytes, 32/64 bit)

[Installation and other instructions](#)

New features in this version: [Windows specific](#), [all platforms](#).

If you want to double-check that the package you have downloaded exactly matches the package distributed by R, you can compare the [md5sum](#) of the .exe to the [true fingerprint](#). You will need a version of md5sum for windows: both [graphical](#) and [command line versions](#) are available.

**Frequently asked questions**

- [How do I install R when using Windows Vista?](#)
- [How do I update packages in my previous version of R?](#)
- [Should I run 32-bit or 64-bit R?](#)

Please see the [R FAQ](#) for general information about R and the [R Windows FAQ](#) for Windows-specific information.

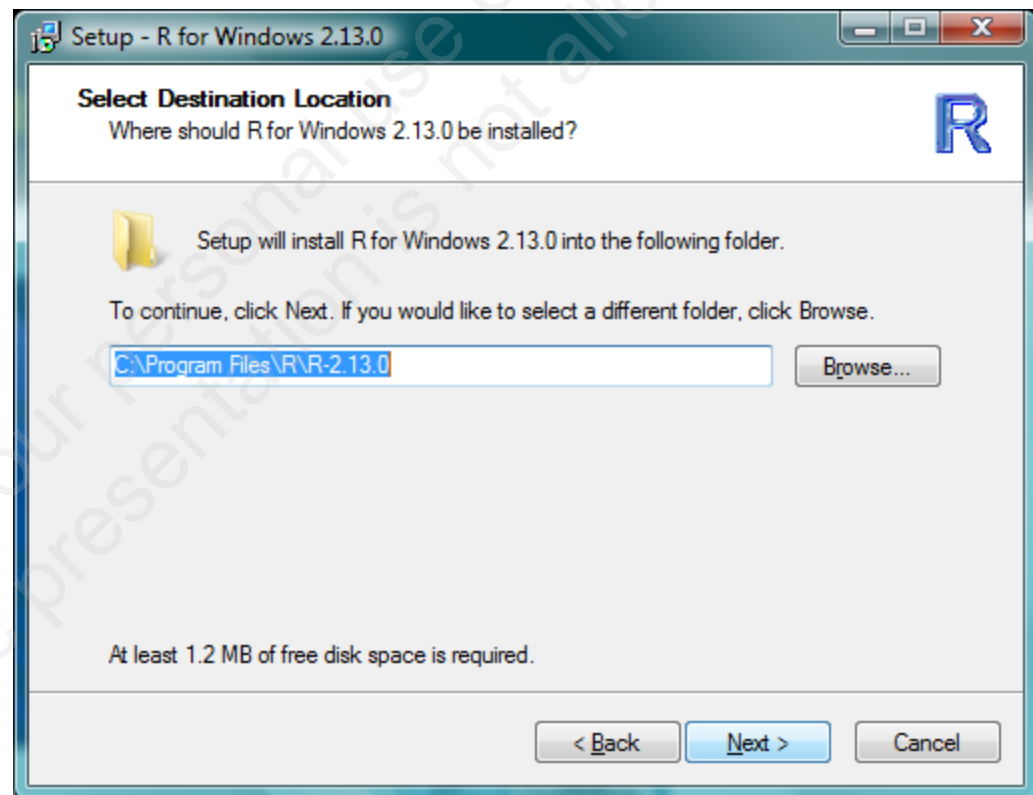
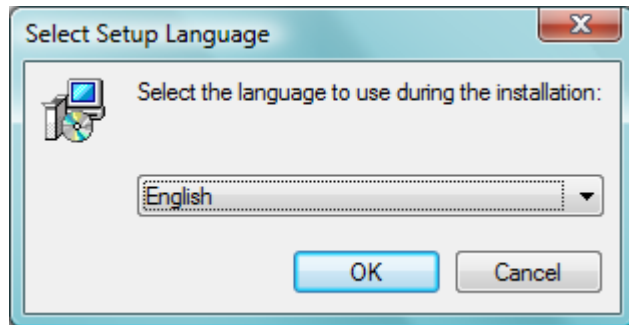
**Other builds**

- Patches to this release are incorporated in the [r-patched snapshot build](#).
- A build of the development version (which will eventually become the next major release of R) is available in the [r-devel snapshot build](#).
- [Previous releases](#)

Note to webmasters: A stable link which will redirect to the current Windows binary release is [<CRAN MIRROR>/bin/windows/base/release.htm](#).

Last change: 2011-04-13, by Duncan Murdoch

# Installing R



... just follow the instructions

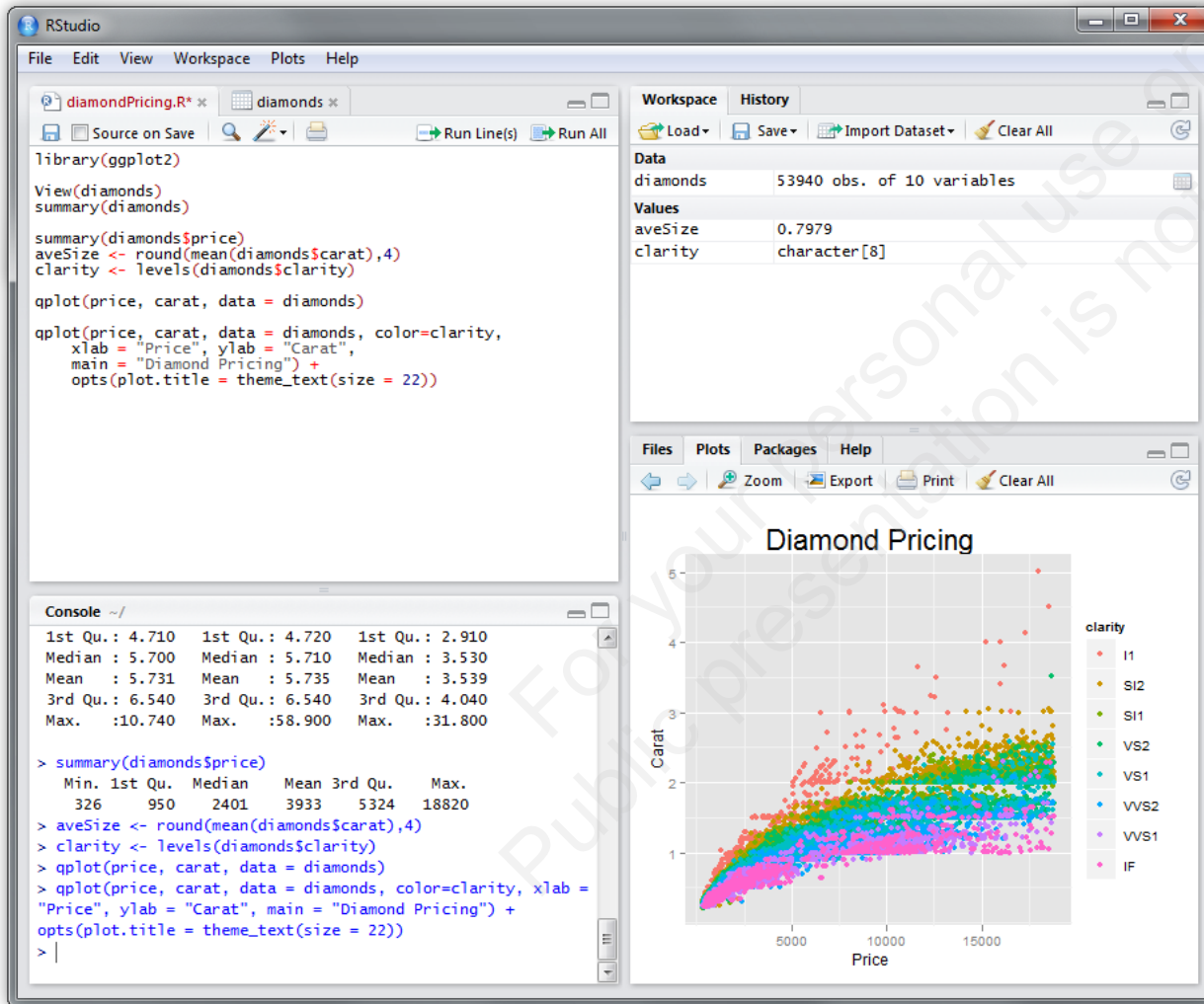
# Examine the base R's GUI



# RStudio: <http://rstudio.org>



- Integrated Development Environment (IDE) for R
- Combines an intuitive user interface with powerful coding tools to help you get the most out of R



# RStudio installation package



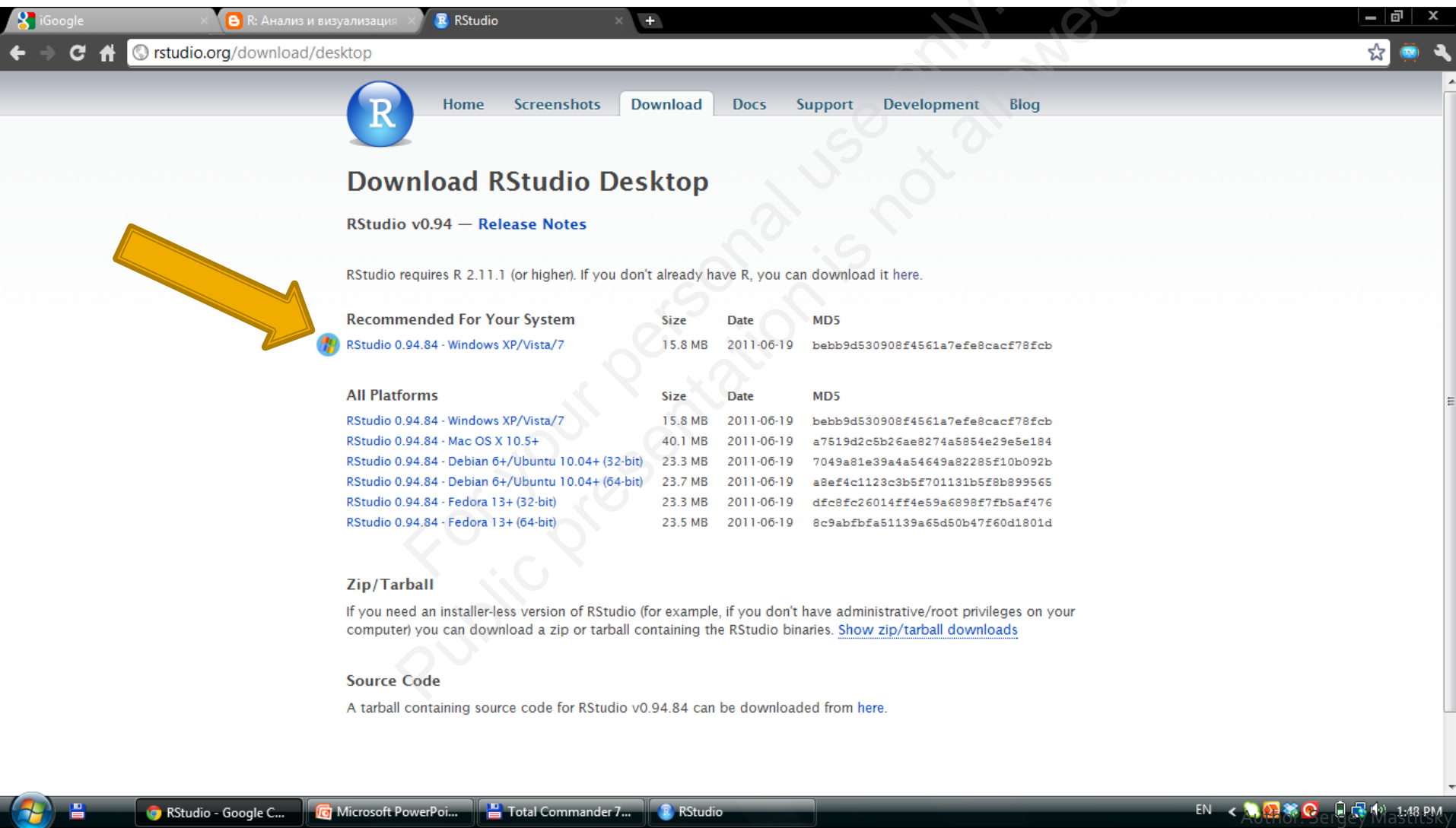
The image shows a screenshot of a web browser displaying the RStudio download page. The browser's address bar shows the URL `rstudio.org/download/`. The page features a navigation menu with links for Home, Screenshots, Download, Docs, Support, Development, and Blog. The main heading is "Download RStudio v0.94".

There are two installation options:

- Desktop:** An icon of a computer monitor with the R logo. The text reads "If you run R on your desktop:" followed by a blue button labeled "Download RStudio Desktop". A large yellow arrow points to this button.
- Server:** An icon of a cloud with the R logo and three server racks below it. The text reads "If you run R on a Linux server and want to enable users to remotely access RStudio using a web browser:" followed by a blue button labeled "Download RStudio Server".

The page also contains a large, faint watermark that reads "For your personal use only. Public presentation is not allowed." The Windows taskbar at the bottom shows several open applications: RStudio - Google C..., Microsoft PowerPoi..., Total Commander 7..., and RStudio. The system tray on the right shows the time as 1:47 PM.

# RStudio installation package




Download RStudio Desktop

RStudio v0.94 — [Release Notes](#)

RStudio requires R 2.11.1 (or higher). If you don't already have R, you can download it [here](#).

**Recommended For Your System**

	Size	Date	MD5
 RStudio 0.94.84 - Windows XP/Vista/7	15.8 MB	2011-06-19	bebb9d530908f4561a7efe8cacf78fcb

**All Platforms**

	Size	Date	MD5
RStudio 0.94.84 - Windows XP/Vista/7	15.8 MB	2011-06-19	bebb9d530908f4561a7efe8cacf78fcb
RStudio 0.94.84 - Mac OS X 10.5+	40.1 MB	2011-06-19	a7519d2c5b26ae8274a5854e29e5e184
RStudio 0.94.84 - Debian 6+/Ubuntu 10.04+ (32-bit)	23.3 MB	2011-06-19	7049a81e39a4a54649a82285f10b092b
RStudio 0.94.84 - Debian 6+/Ubuntu 10.04+ (64-bit)	23.7 MB	2011-06-19	a8ef4c1123c3b5f701131b5f8b899565
RStudio 0.94.84 - Fedora 13+ (32-bit)	23.3 MB	2011-06-19	dfc8fc26014ff4e59a6898f7fb5af476
RStudio 0.94.84 - Fedora 13+ (64-bit)	23.5 MB	2011-06-19	8c9abfbfa51139a65d50b47f60d1801d

**Zip/Tarball**

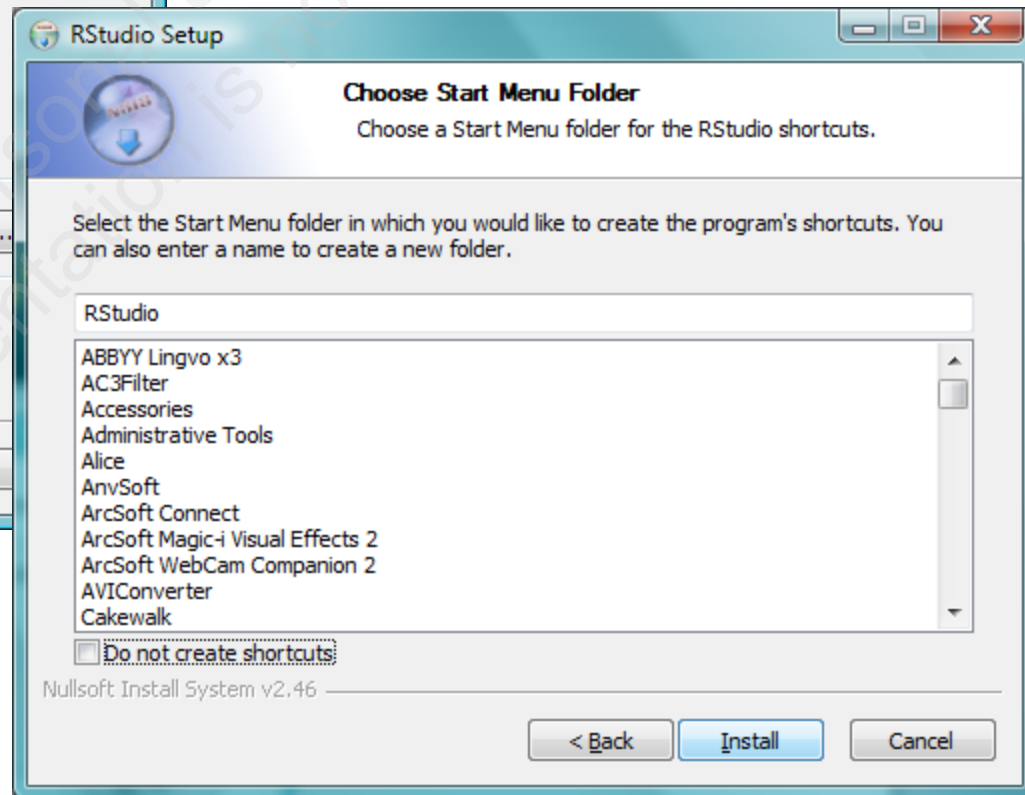
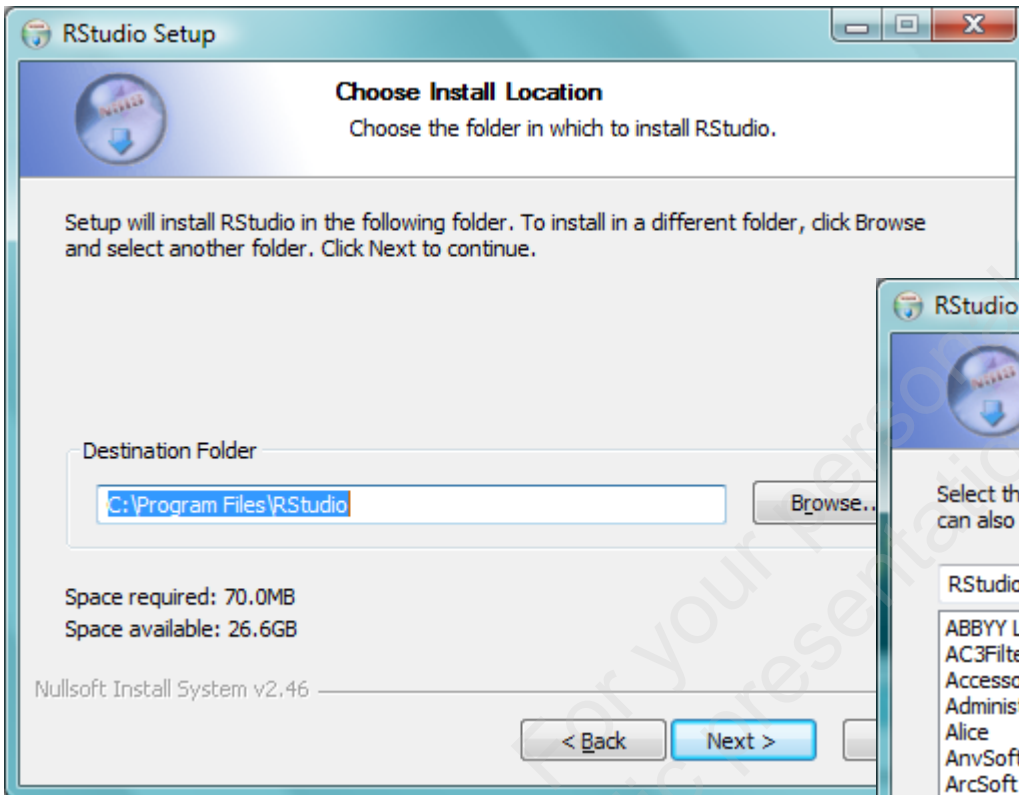
If you need an installer-less version of RStudio (for example, if you don't have administrative/root privileges on your computer) you can download a zip or tarball containing the RStudio binaries. [Show zip/tarball downloads](#)

**Source Code**

A tarball containing source code for RStudio v0.94.84 can be downloaded from [here](#).



# Installing RStudio



# Work window of RStudio

The image shows the RStudio interface with three red callout boxes highlighting key components:

- Script editor: develop you code here**: Points to the main editor window where R code is written.
- R console: here you talk with R directly, by entering commands**: Points to the console window where R commands are executed and output is displayed.
- Service window 1**: Points to the **Packages** panel, which lists installed and available R packages.
- Service window 2**: Points to the **Workspace** panel, which shows the current environment and workspace.

The console window displays the following text:

```
R version 2.13.0 (2011-04-13)
Copyright (C) 2011 The R Foundation for Statistical Computing
ISBN 3-900051-07-0
Platform: x86_64-pc-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.

Natural language support but running in an English locale

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.

>
```

# Exercises

- Go to the R Project website and get yourself familiar with its main sections
- From the subsection “Documentation/Manuals”, download PDFs of the “[An Introduction to R](#)” and “[R Data Import/Export](#)” – for your further reading (“must have”!)
- Pay attention to other subsections of “Documentation”