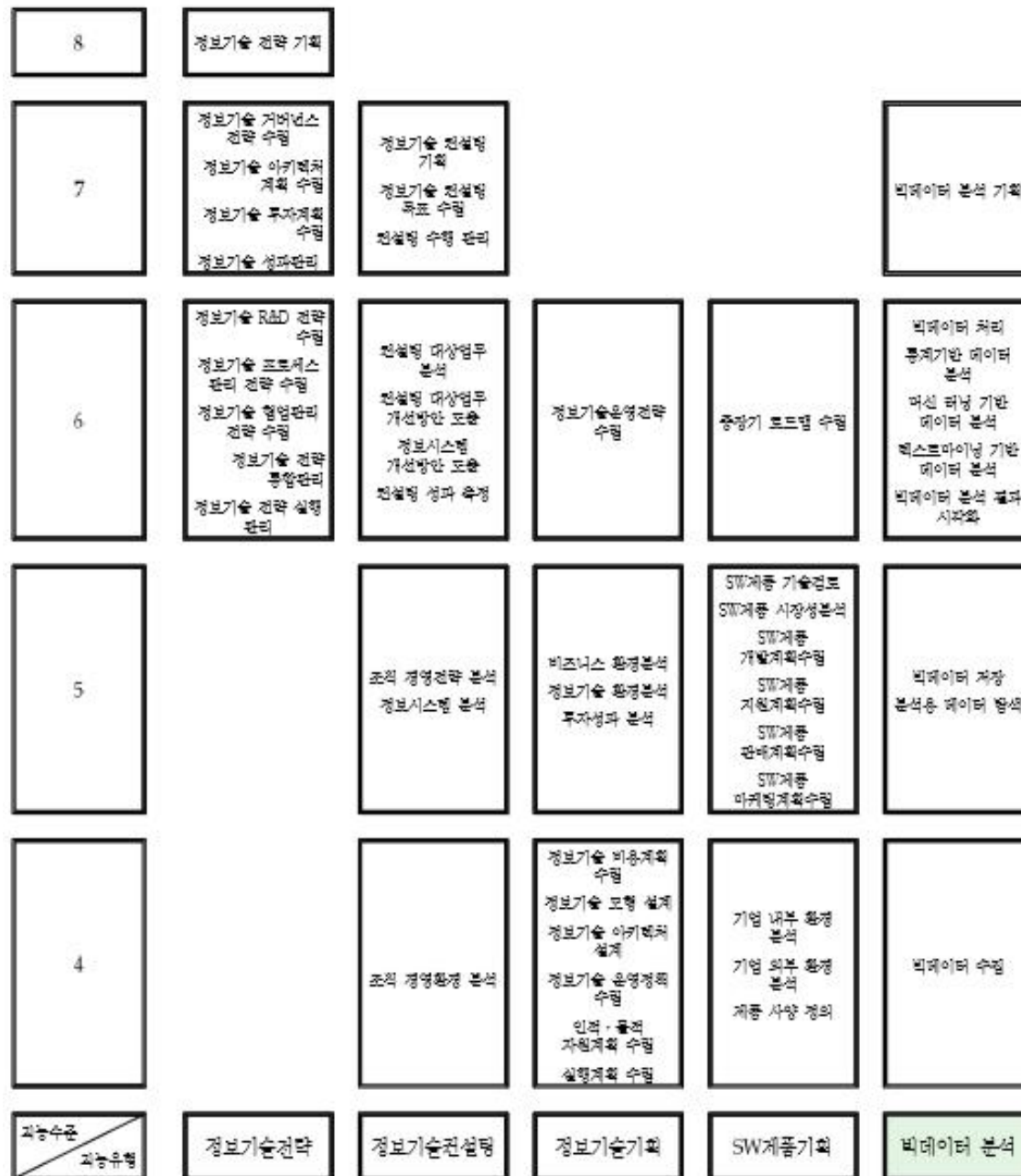


빅데이터 분석 체계도 (NCS 기준)



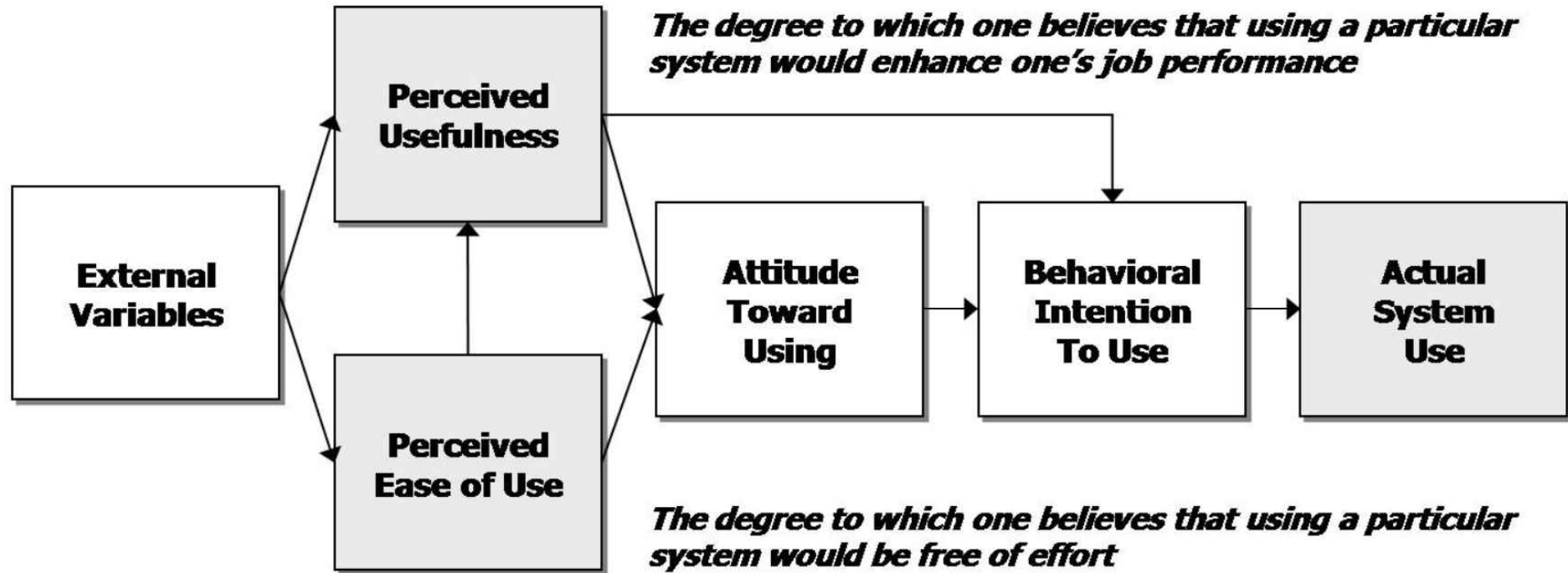
빅데이터 모델링

1. 가설설정: 수집된 내·외부 데이터 및 정형·비정형 데이터를 활용하여 분석 목적에 따라 가설을 설정하고 필요한 데이터셋을 편성하여 통계기반 데이터 분석 모델을 만들고 평가하는 능력필요.
2. 빅데이터 모델개발: 설정된 가설을 검증하기 위하여 분석모형을 설계, 모델을 완성도를 평가하기 위한 성능평가 기준을 설정,모델에 적용하는 독립변수와 종속변수의 데이터 유형에 따라 다양한 분석기법을 적용,주어진 데이터 변수 추가하여 모델에서 필요로 되어지는 파생변수(Drived Variable)를 생성하여 모델링에 활용.
3. 빅데이터 모델 평가 검증하기 :구현된 모델은 모델 운영의 안정성과 적합성을 평가, 설정되어 있는 모델의 평가 절차 및 기준에 따라 구현된 모델을 평가, 모델 평가를 위한 데이터셋은 훈련 데이터셋과 구분되어 관리되어 평가의 객관성을 확보, 모델 평가의 전과정을 절차에 따라 진행되었음을 증명하고 모델 평가결과 를 리포팅 해야함

빅데이터 모델링 고려 사항

1. 정형/비정형 데이터를 분석 목적에 따라 가설을 설정하고 분석 모델을 만들어 평가하는 업무 적용.
2. 예측 모형은 회귀분석, [SVM](#), 의사결정 Tree, 분류모형은 의사결정 Tree 신경망, 로지스틱 회귀분석, [KNN](#).
3. 클러스터링 모형은 [K-Means](#) 등을 포함한다.
4. 데이터분석 시각화를 활용한다.
5. 분석도구는 오픈소스 소프트웨어인 R과 상업용 패키지를 활용한다.

데이터분석 모델링 (정보기술수용모델을 기본으로 한 예측)



직각된 유용성과 지각된 사용용이성이 태도와 의도에 직접적인 영향을 줌으로 행동에 영향을 미치고 있음

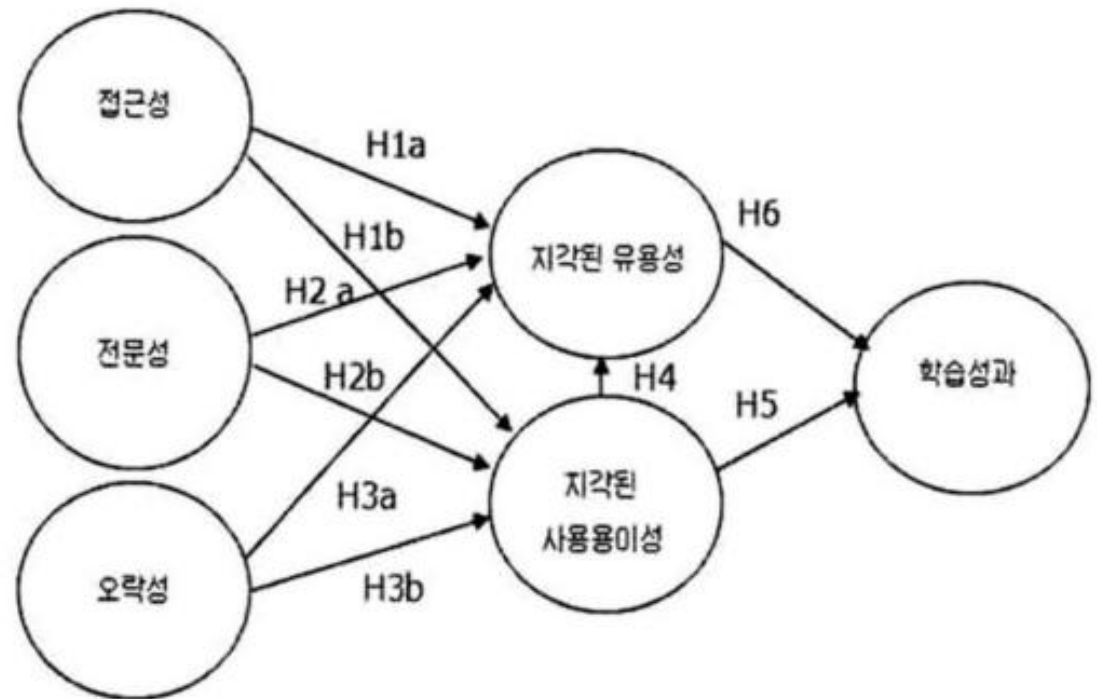
데이터분석 모델링 예)



지각된 유용성과 지각된 사용용이성이 웹 기반 원격교육의 학습 성과에 미치는 영향
- 국내 고등학생을 중심으로

The effect of Perceived usefulness and Perceived easy of use on Web Based Distance
High School Students

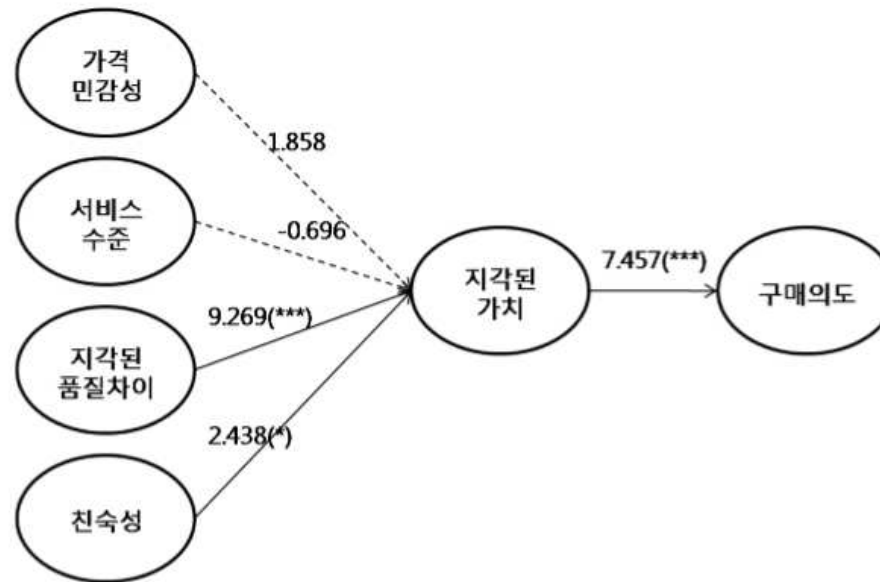
저자 (Authors)	정화민, 조원길 Hwa-Min Jeong, Won-Gil Cho
출처 (Source)	e-비즈니스연구 12(2) , 2011.06, 255-272 (18 pages) The e-Business Studies 12(2) , 2011.06, 255-272 (18 pages)
발행처 (Publisher)	국제e-비즈니스학회 Global E-Business Association
URL	http://www.dbpia.co.kr/Article/NODE01769045
APA Style	정화민, 조원길 (2011). 지각된 유용성과 지각된 사용용이성이 웹 기반 원격교육의 학습 성과에 미치는 영향. e-비즈니스연구, 12(2), 255-272.
이용정보 (Accessed)	서강대학교 163.***.1.208 2018/09/20 17:35 (KST)



[그림 1] 본 연구의 실증 연구 모형

학습성과를 예측하는 모델링

데이터분석 모델링 예: PB제품 재구매의도



〈그림 2〉 가설검증을 위한 경로분석

〈표 10〉 경로분석

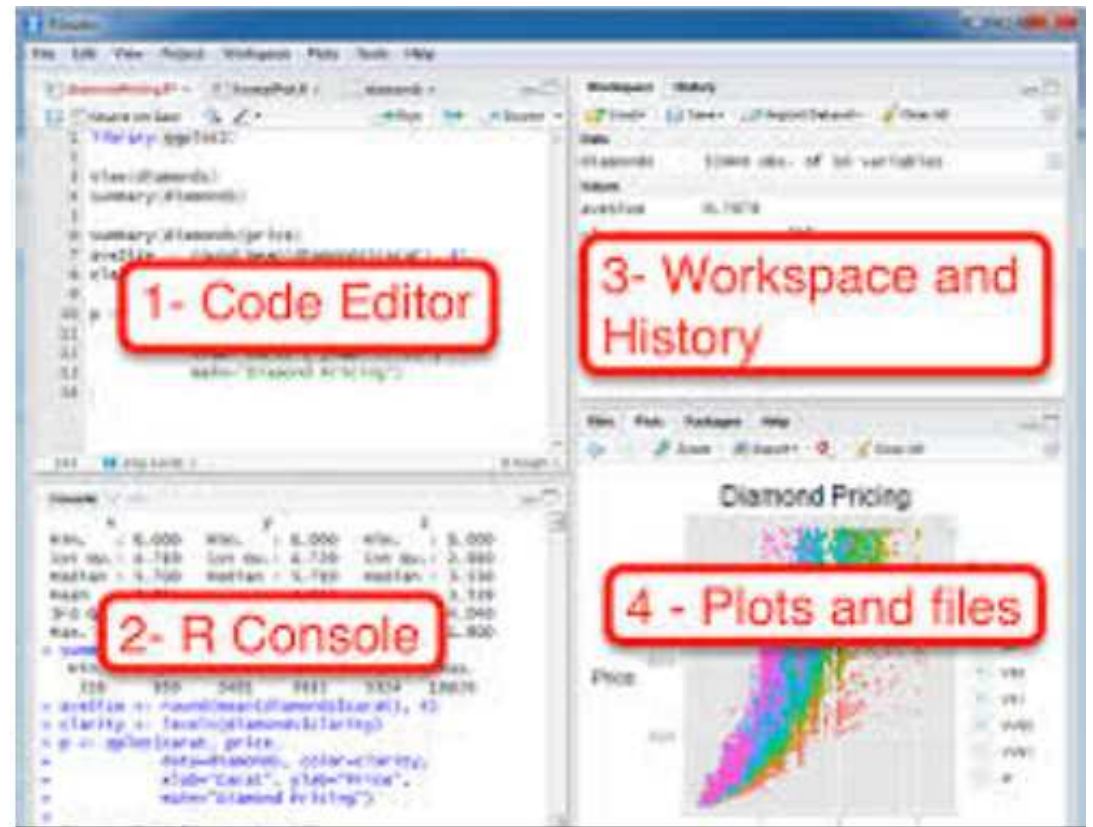
가설	가설경로			Estimate	S.E.	C.R.	P	결과
H1	지각된_가치	<-	가격_민감성	0.067	0.036	1.858	0.063	기각
H2	지각된_가치	<-	서비스_수준	-0.038	0.055	-0.696	0.487	기각
H3	지각된_가치	<-	지각된_품질	0.746	0.081	9.269	***	채택
H4	지각된_가치	<-	제품 친숙성	0.096	0.039	2.438	0.015	채택
H5	구매_의도	<-	지각된_가치	0.959	0.129	7.457	***	채택

데이터분석 모델링 분석결과

연구가설	경로			계수값	표준오차	CR값	P값	결과
H1a	유용성	<---	접근성	0.045	0.061	0.746	0.456	기각
H1b	용이성	<---	접근성	-0.025	0.101	-0.25	0.803	기각
H2a	유용성	<---	전문성	0.142	0.045	3.185	0.001	채택
H2b	용이성	<---	전문성	-0.139	0.071	-1.96	0.05	채택
H3a	유용성	<---	오락성	0.104	0.05	2.071	0.038	채택
H3b	용이성	<---	오락성	0.327	0.082	4.008	0.000	채택
H4	유용성	<---	용이성	0.264	0.047	5.682	0.000	채택
H5	학업성취	<---	용이성	0.26	0.06	4.37	0.000	채택
H6	학업성취	<---	유용성	0.328	0.103	3.176	0.001	채택

전문성은 유용성에 0.142 만큼 긍정적 영향
 유용성은 학업성취에 0.328만큼 긍정적 영향을 주고 있음

데이터 분석 R 프로그래밍 (R, R 스튜디오 설치)



R 설치

r studio 사용법 pdf



전체

이미지

동영상

뉴스

지도

더보기

설정

도구

검색결과 약 32,200개 중 3페이지 (0.32초)

R 설치, 기초 사용법 – Project Elite

projectl33t.xyz/archives/244 ▼

2016. 10. 17. - 또한 **R-Studio**는 관리자 권한으로 실행되도록 해주세요. ... 뿐만 아니라 출력된 시각화 자료를 이미지나 **pdf** 파일로 저장하는 기능도 있습니다.

R markdown 사용법 - Amazon AWS

https://rstudio-pubs-static.s3.amazonaws.com/369288_2e88c27061494355977e770ece7...

R과 **RStudio** 설치방법: eclass의 동영상 참고; 라이브러리 설치 ... knitr을 이용하면 R코드와 설명을 html, latex, **pdf** 등 다양한 형태로 변환 가능; R script(확장자는 .

R 설치 후 PC 상에서의 실행

The screenshot displays the RStudio environment with the following components:

- Source Editor:** Contains an R script with the following code:

```
1 install.packages(pkgs, lib, repos = getOption("repos"), contriburl = contrib.url(repos, type),
2 ##
3 # 패키지 설치
4 install.packages("rJava")
5 install.packages("memoise")
6 install.packages("konlp")
7 install.packages("dplyr")
8
9 # 패키지 로드
10 library(konlp)
11 library(dplyr)
12
13 # java 폴더 경로 설정
14 Sys.setenv(JAVA_HOME="C:/Program Files/Java/jre1.8.0_111/")
15
16 useNIADic()
17
18 # 데이터 불러오기
19 txt <- readLines("hiphop.txt")
20 head(txt)
21
22 install.packages("stringr")
```
- Console:** Shows the R version and platform information:

```
R version 3.5.1 (2018-07-02) -- "Feather Spray"
Copyright (C) 2018 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 D:/음악/음악/대원/대원고/음악/대원고/대원고/Data/.RData]
```
- Environment:** Displays the loaded packages and the current environment:

```
useNIADic()
library(dplyr)

##
# 패키지 설치
install.packages("rJava")
install.packages("memoise")
install.packages("memoise")
library(dplyr)
# 데이터 불러오기
txt <- readLines("hiphop.txt")
# 데이터 불러오기
txt <- readLines("hiphop.txt")
head(txt)
head(txt)
# 데이터 불러오기
txt <- readLines("hiphop.txt")
head(txt)
```
- Files:** Shows the installed packages and their versions:

Name	Description	Version
System Library		
<input type="checkbox"/> assertthat	Easy Pre and Post Assertions	0.2.0
<input type="checkbox"/> BH	Boost C++ Header Files	1.66.0-1
<input type="checkbox"/> bindr	Parametrized Active Bindings	0.1.1
<input type="checkbox"/> bindrcpp	An 'Rcpp' Interface to Active Bindings	0.2.2
<input type="checkbox"/> bit	A Class for Vectors of 1-Bit Booleans	1.1-14
<input type="checkbox"/> bit64	A S3 Class for Vectors of 64bit Integers	0.9-7
<input type="checkbox"/> blob	A Simple S3 Class for Representing Vectors of Binary Data ('BLOBs')	1.1.1
<input type="checkbox"/> boot	Bootstrap Functions (Originally by Angelo Canty for S)	1.3-20
<input type="checkbox"/> class	Functions for Classification	7.3-14
<input type="checkbox"/> cli	Helpers for Developing Command Line Interfaces	1.0.0
<input type="checkbox"/> cluster	"Finding Groups in Data": Cluster Analysis Extended Rousseeuw et al.	2.0.7-1
<input type="checkbox"/> codetools	Code Analysis Tools for R	0.2-15
<input type="checkbox"/> colorspace	Color Space Manipulation	1.3-2
<input type="checkbox"/> compiler	The R Compiler Package	3.5.1
<input type="checkbox"/> crayon	Colored Terminal Output	1.3.4
<input type="checkbox"/> curl	A Modern and Flexible Web Client for R	3.2
<input type="checkbox"/> datasets	The R Datasets Package	3.5.1
<input type="checkbox"/> DBI	R Database Interface	1.0.0
<input type="checkbox"/> devtools	Tools to Make Developing R Packages Easier	1.13.6
<input type="checkbox"/> dichromat	Color Schemes for Dichromats	2.0-0
<input type="checkbox"/> digest	Create Compact Hash Digests of R Objects	0.6.15
<input type="checkbox"/> dplyr	A Grammar of Data Manipulation	0.7.6
<input type="checkbox"/> fansi	ANSI Control Sequence Aware String Functions	0.2.3
<input type="checkbox"/> foreign	Read Data Stored by 'Minitab', 'S', 'SAS', 'SPSS', 'Stata', 'Systat', 'Weka', 'dBase', ...	0.8-70