

고고학 자료 통계분석

Week 9 : 주성분분석

숭실대학교 사학과 석사과정
주 찬 혁

계획

주차	제목	내용
1	Intro	소개, R 설치
2	기초 통계(1)	모집단과 표본, 기술통계량
3	기초 통계(2)	변수의 종류, 가설과 검정, 오류, 분석절차
4	전치과	데이터 전치과
5	사각화	다양한 종류의 그래프
6	검정과 상관분석	검정, 상관분석
7	회귀분석	선형회귀, 다중선형회귀, 로지스틱회귀
8	군집분석과 판별분석	K-means, LDA
9	주성분분석	PCA

복습

- 군집분석이 무엇인지 안다.
- 덴드로그램으로 계층적 군집분석 결과를 시각화할 수 있다.
- R로 K-means 군집분석을 시행할 수 있다.
- 판별분석이 무엇인지 안다.
- R로 선형판별분석을 시행할 수 있다.
- 군집분석과 판별분석의 차이점에 대해 안다.

주성분분석(Principal Component Analysis)

- 본래 데이터의 분포를 최대한 보존시키면서 고차원을 저차원으로 변환하는 분석방법
- 가장 보편적인 차원축소 방법 중 하나
- 데이터를 가장 잘 표현하는 조합을 찾고 이를 통해 차원을 축소함
- 기존 변수들을 반영하는 Principal Component를 만들어 대리지표로 활용
- 통상적으로 proportion of variance의 값이 합해서 0.9(90%) 이상이 되는 PC 개수를 선정하고 시각화함

주성분분석(PCA)

> 모델명 <- prcomp(데이터, center = T, scale. = T)

> summary(모델명)

> 플롯명 <- ggbiplot(모델명, 옵션 지정)

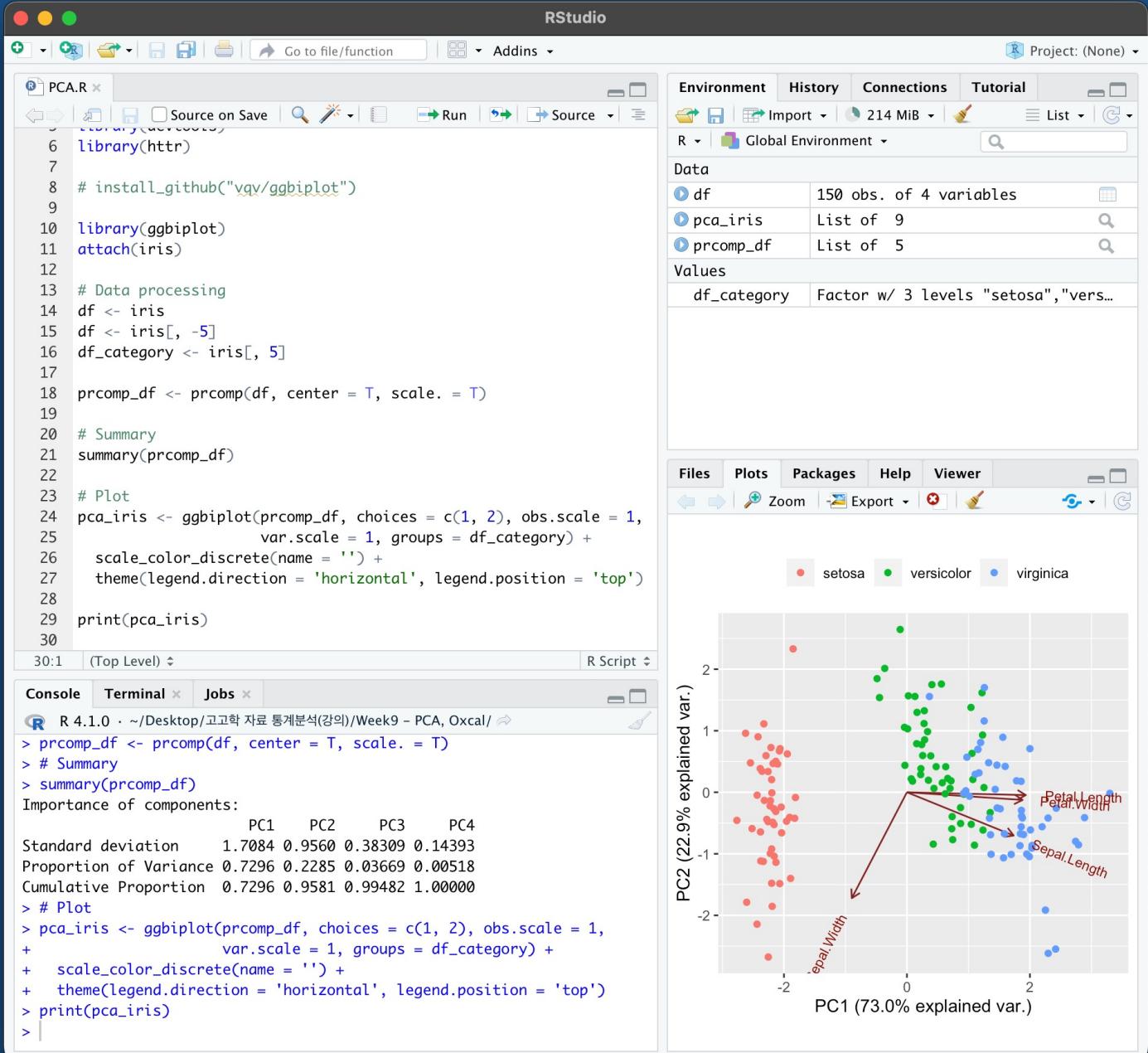
> print(플롯명)

- prcomp 명령어로 PCA 시행

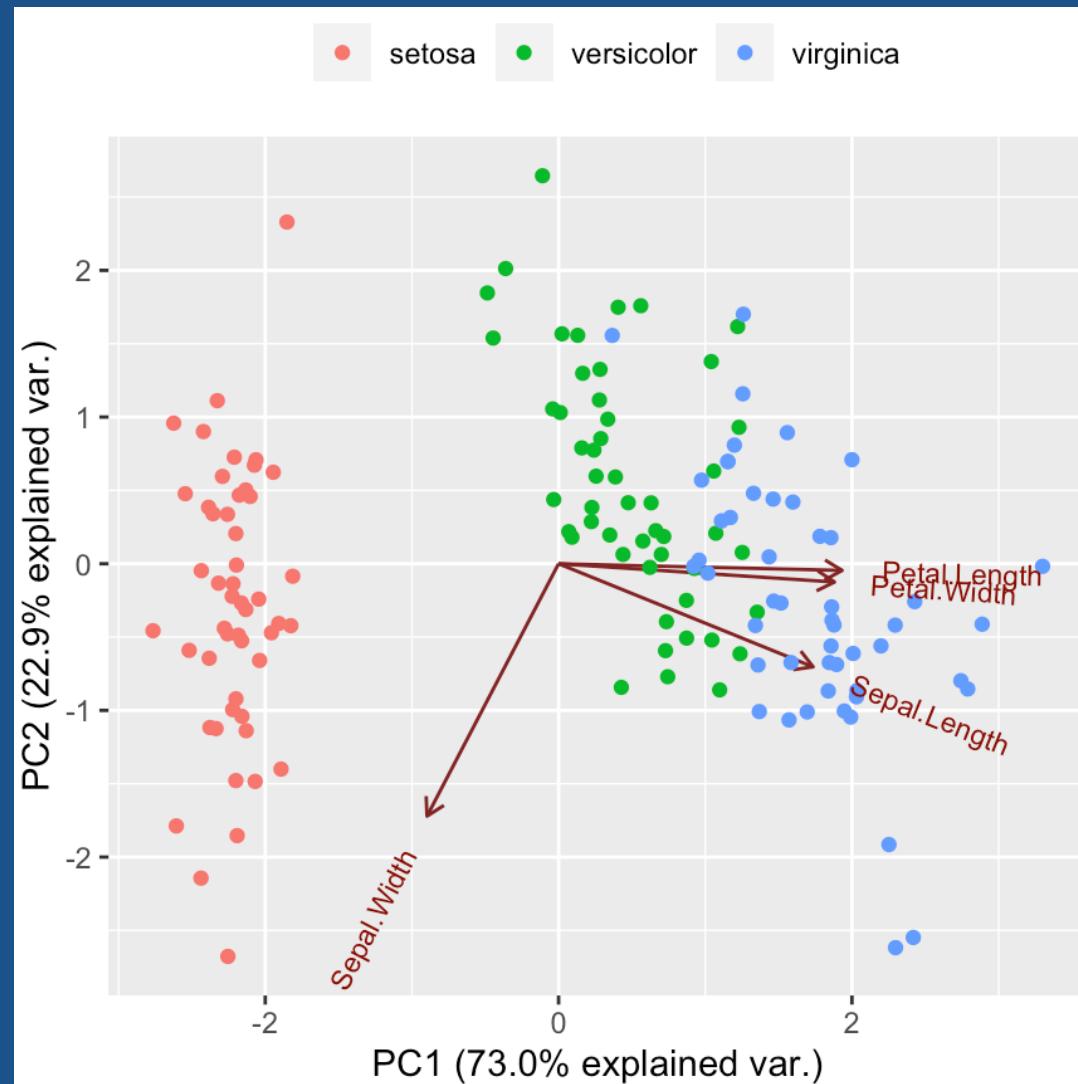
- summary로 모델 지표 확인

- ggbiplot으로 시각화

*데이터 스케일에 민감하기 때문에 센터링과 정규화는 반드시 시행해줘야함



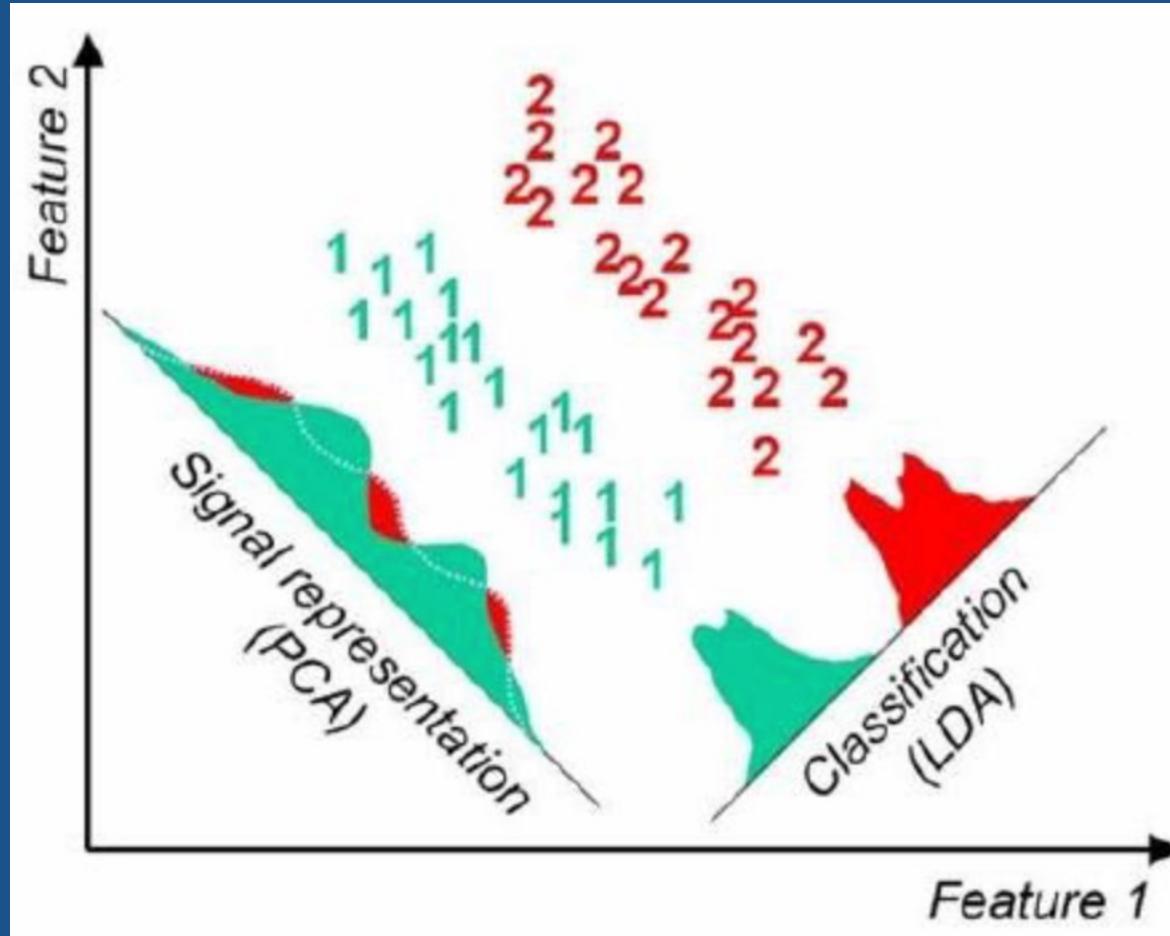
주성분분석(PCA)



주성분분석(Clustering)

- 주성분분석은 **센터링**과 **정규화**가 이루어졌다는 가정 위에서 시행됨.
- **센터링(Centering)** : 각 **변수의 평균을 0으로** 만드는 것
 - 다중공선성을 줄일 수 있음
- **정규화(Normalization)** : **변수들의 척도를 통일해주는 것**
 - 다른 단위가진 변수는 원칙적으로 비교할 수 없음
 - 모든 값을 0~1로 변환

주성분분석과 판별분석의 차이



고고학 자료 통계분석

부록 1 : 다양한 R 패키지

숭실대학교 사학과 석사과정
주 찬 혁

Computational Archaeology

- 컴퓨터를 활용한 대용량 연산을 통해 다양한 고고학적 문제를 해결
- 컴퓨터를 통해 GIS, 통계, 머신러닝 등의 방법을 사용하는 경우
- 1주차에 알아보았듯, 고고학 연구와 관련된 패키지들은 대부분 R로 구현되어 있음
- 주로 영국, 독일, 미국을 중심으로 응용과 방법론 개량이 이루어지고 있음

Ben Marwick의 ctv-Archaeology

- University of washington의 교수 Ben Marwick이 개설해놓은 Github Repository로 고고학의 여러 분야에 응용되는 R 패키지들을 정리, 요약함
- 오픈소스 프로젝트로 운영되어, 각 패키지 개발자들이 소개 내용을 직접 업로드
- 소개되어 있는 패키지들을 일괄적으로 설치하는 기능을 제공
- Link : <https://github.com/benmarwick/ctv-archaeology>



Ben Marwick

고고학 자료 통계분석

부록 2 : Oxcal

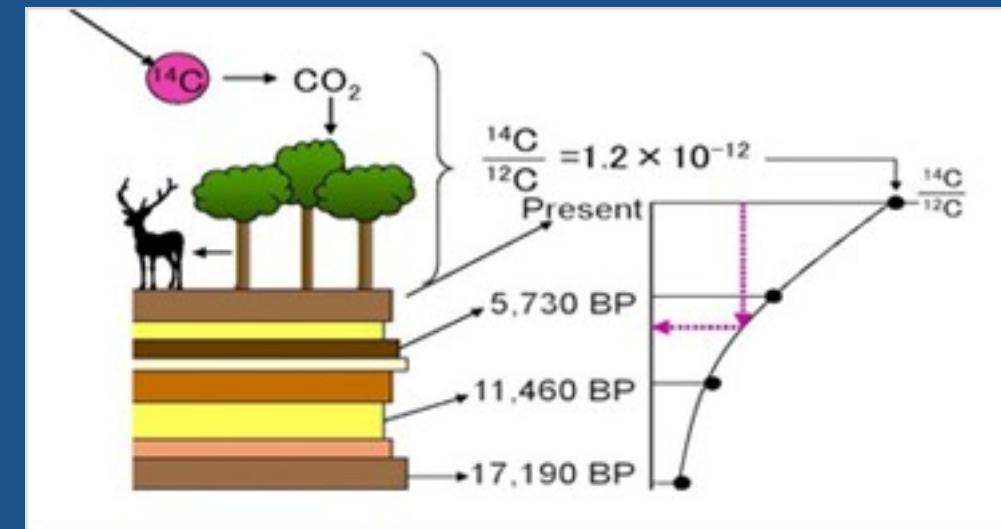
숭실대학교 사학과 석사과정
주 찬 혁

방사성탄소연대측정법?

- 절대연대측정법 중 하나
- 탄소(C)의 동위원소인 C14의 비율을 측정하여 시료의 연대를 추정하는 연대측정법입니다.
- 정확히는 시료에 잔존하는 C12, C13과 C14의 비율을 비교하여 산출된 시료의 생명활동이 멈춘 시기를 의미

방사성탄소연대측정법?

- 대기 중에 일정한 비율로 분포하고 있는 C14는 식물의 광합성과 동물의 호흡을 통해 개체의 체내에 누적, 누적되는 비율은 대기 중의 비율과 동일
- 이후 개체가 죽음으로 인해 광합성과 호흡을 멈추게 되면 안정동위원소인 C12, C13의 수는 유지되지만, 불안정동위원소인 C14는 붕괴.
- 이 때 C14의 반감기 일정하기에, 이를 통해서 시료의 죽음으로부터 경과된 시간을 추정할 수 있음.



방사성탄소연대측정법의 문제점

1. 대기 중의 C14 비율이 변함

- : 방사성탄소연대측정법은 C14 비율이 일정하다는 전제 위에서 작동함
- : 그러나 산업화와 핵실험 등의 영향으로 인해 대기중의 C14비율이 변함

2. 고목효과

- : 나무가 벌목된 연대와 실제 탄화된 시점 사이의 간극이 있을 수 있음

해양 리저브 효과, 할슈타트 평탄면 등 ...

할슈타트 평탄면(Hallstatt Plateau)

- 2450BP 인근으로 보정곡선이 일괄적으로 평평한 영역
- 이 구간에 해당 측정치들을 보정할 경우, 전부 B.C. 800-400로 보정됨
- 위글매칭 등의 방법으로 부분적으로 극복

수준연대보정법이란?

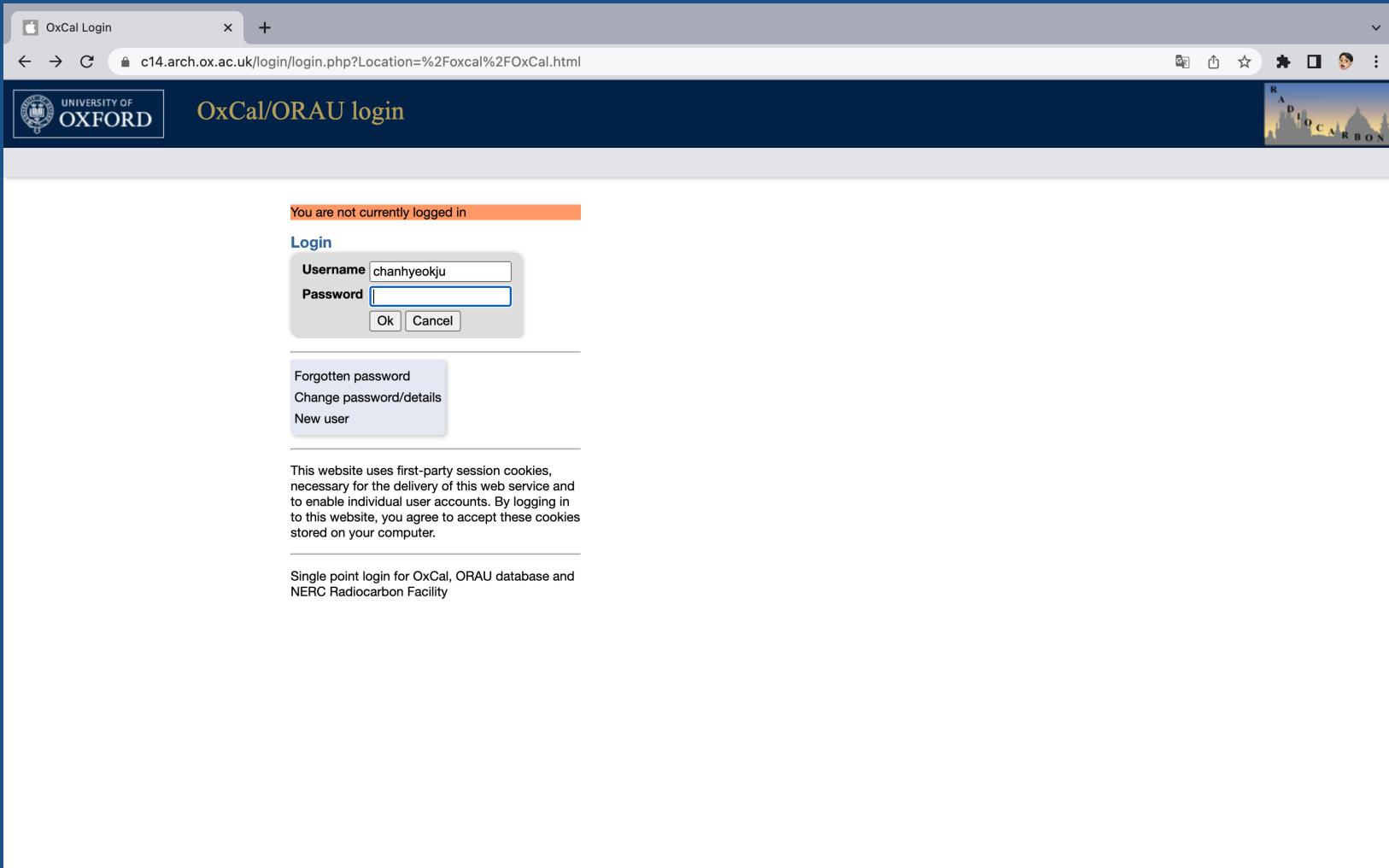
- 연륜연대결정법을 응용하여 만들어진 보정곡선을 통해
방사성탄소연대측정치를 보정하는 방법
- 자세한건 이창희, 방사성탄소연대측정법의 이해와 활용 참고
 - Link : <http://edu.kaah.kr/board/sub1>

OxCat

- 방사성탄소연대보정 프로그램
- University of Oxford에서 개발 및 배포
- Oxford는 연대보정에 가장 보편적으로 활용되는 IntCal 보정곡선을 만들고 있음(최근 2020년에 IntCal20을 배포)
- 방사성탄소연대를 보정하는 다른 프로그램(R의 Bchron, rcarbon, ADMUR 등)들도 있지만, OxCat은 일종의 표준역할로 사용됨.



OxCAL - 가입 및 로그인



OxCal - 최초 화면

File -> New

The screenshot shows the OxCal 4.4 web interface. On the left, a sidebar displays the 'File' menu with options: New, Open, Close, Save as, Manager, Reset, Change password, Logout, Print..., and Export... A red box highlights the 'File -> New' option. The main content area is titled 'OxCal 4.4 Manual'. It features a search bar at the top right. Below the title, there's a list of links: Introduction, Licence, Installation, Input of information, Analysis, Calibration Curves, Viewing output, Command reference, Program development history, References and bibliography, and Acknowledgements. Further down, it shows the Web interface build number (133), last update date (24/11/2021), and copyright notice ((c) Christopher Bronk Ramsey 2021). A section titled 'References' lists various academic papers and publications. At the bottom, there's a note about citation requirements and a brief introduction to the program.

OxCal 4.4 Manual

File View Format Options Data Help

New
Open
Close
Save as

Manager
Reset
Change password
Logout

Print...
Export...

Introduction
Licence
Installation
Input of information
Analysis
Calibration Curves
Viewing output
Command reference
Program development history
References and bibliography
Acknowledgements

Web interface build number: 133
Last Updated: 24/11/2021
(c) Christopher Bronk Ramsey 2021

References

- Bronk Ramsey 1994 (first release notice)
- Bronk Ramsey 1995 (main introduction to program)
- Bronk Ramsey 1998 (approach and plans)
- Bronk Ramsey 2001 (developments 1995-2001)
- Bronk Ramsey et al. 2001 (D_Sequence method and tests of other methods)
- Bronk Ramsey 2008 (deposition models) [pre-print]
- Bronk Ramsey 2009a (new features of OxCal4) [pre-print]
- Bronk Ramsey 2009b (outlier analysis) [pre-print]
- Bronk Ramsey et al. 2010
- Lee and Bronk Ramsey 2012 (trapezoidal models)
- Radiocarbon Vol 55(4) (IntCal13 special issue)
- Bronk Ramsey and Lee 2013 (variable k in P_Sequence and other developments)
- Bronk Ramsey 2017 KDE plots and models
- Radiocarbon IntCal20 special issue
 - Statistical approach (Heaton et al. 2020)
 - IntCal20: Northern Hemisphere (Reimer et al. 2020)
 - SHCal20: Southern Hemisphere (Hogg et al. 2020)
 - Marine20 - Modelled ocean average (Heaton et al. 2020)

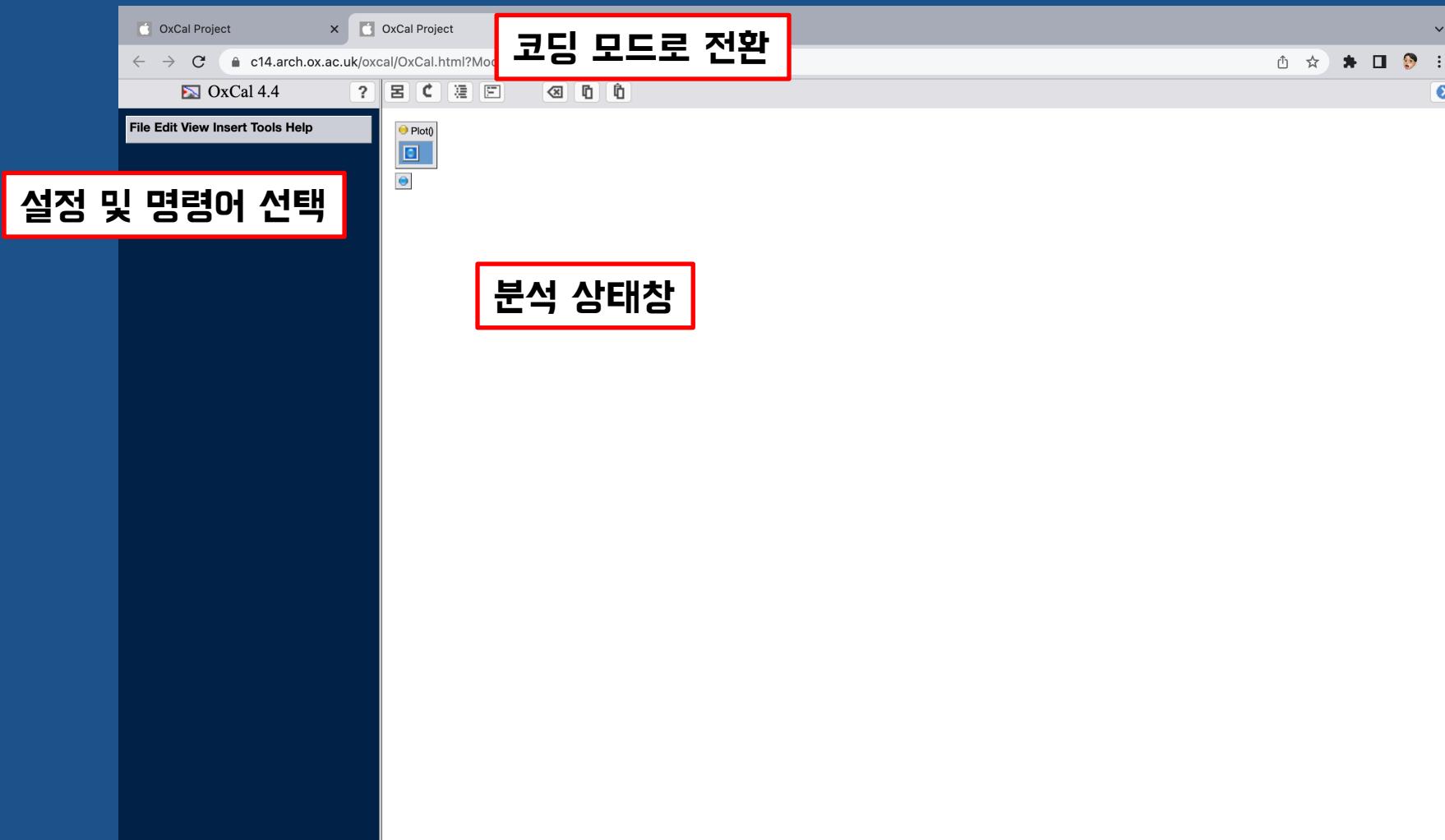
If you use this program, you should quote the reference for the calibration curve used, the version of OxCal (with any non-standard options set) and the reference Bronk Ramsey 2009a. If you are wiggle-matching tree-ring sequences you should quote Bronk Ramsey et al. 2001. For deposition models Bronk Ramsey 2008 should be referenced, together with Bronk Ramsey and Lee 2013 for variable k, and for outlier analysis Bronk Ramsey 2009b. For use of trapezoidal priors please quote Lee and Bronk Ramsey 2012. Finally for KDE plots and models use the reference for Bronk Ramsey 2017.

Introduction

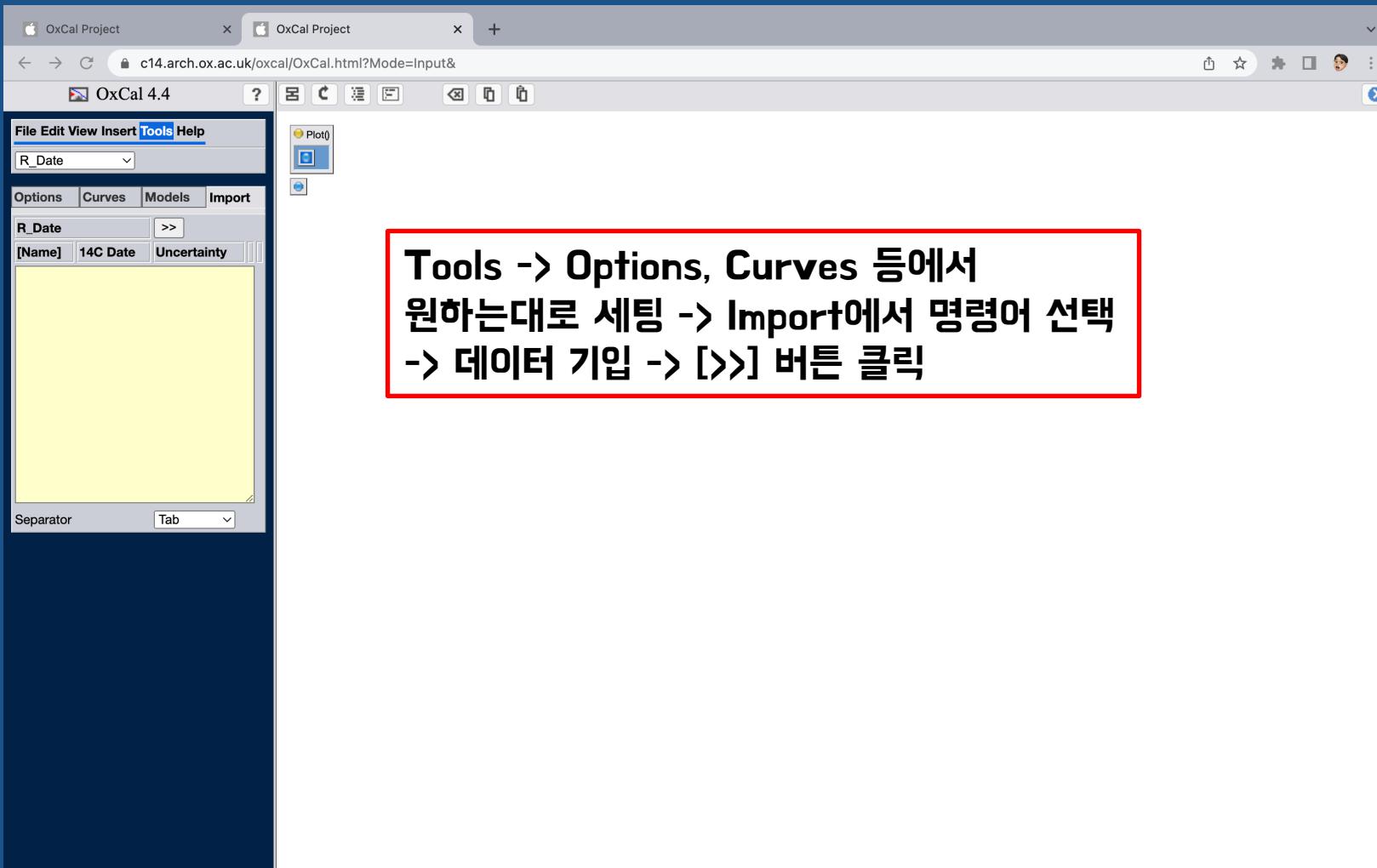
OxCal is a program designed for the analysis of chronological information.

IntCal 20

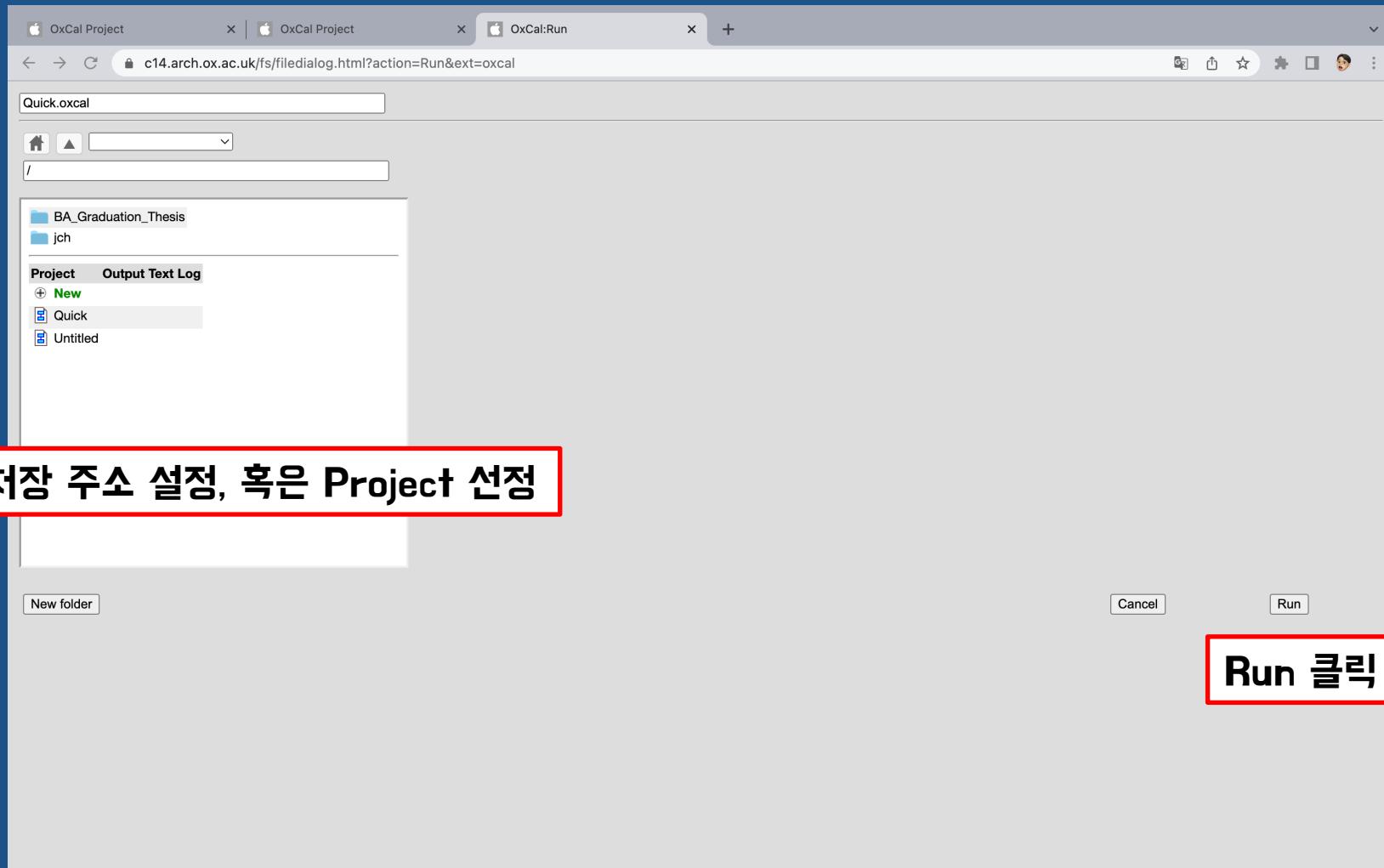
OxCal - 초기 화면



OxCal - 초기 화면



OxCal - 저장 화면



0xCal - 다양한 명령어

- **R-Date** : 미보정 연대를 보정
- **R-Combine** : 동일 시료 혹은 동일 지점에서 발견된 두 개 이상의 시료의 미보정 연대를 하나의 연대로 산출
- **SUM, R-outlier, R-Simulate** 등...

R-Date

1. R-Date 선택

4. 분석 상태창에서 확인

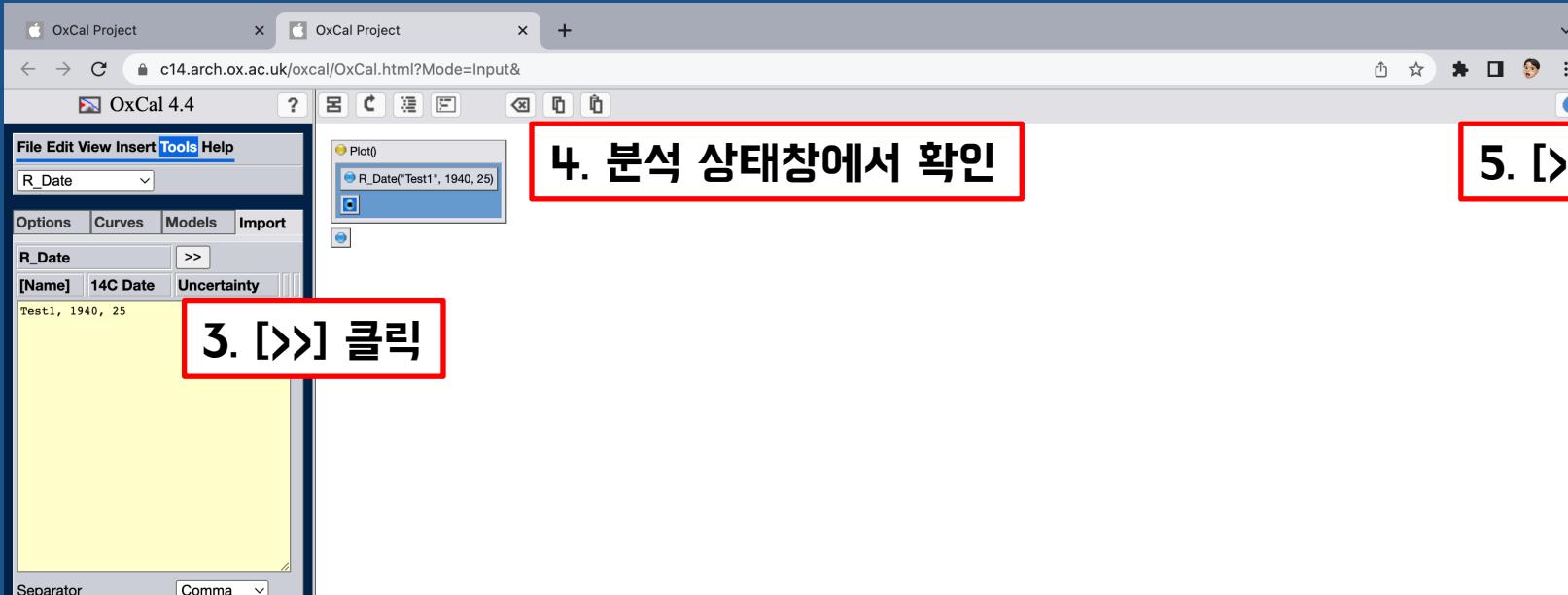
5. [>] 클릭

3. [>>] 클릭

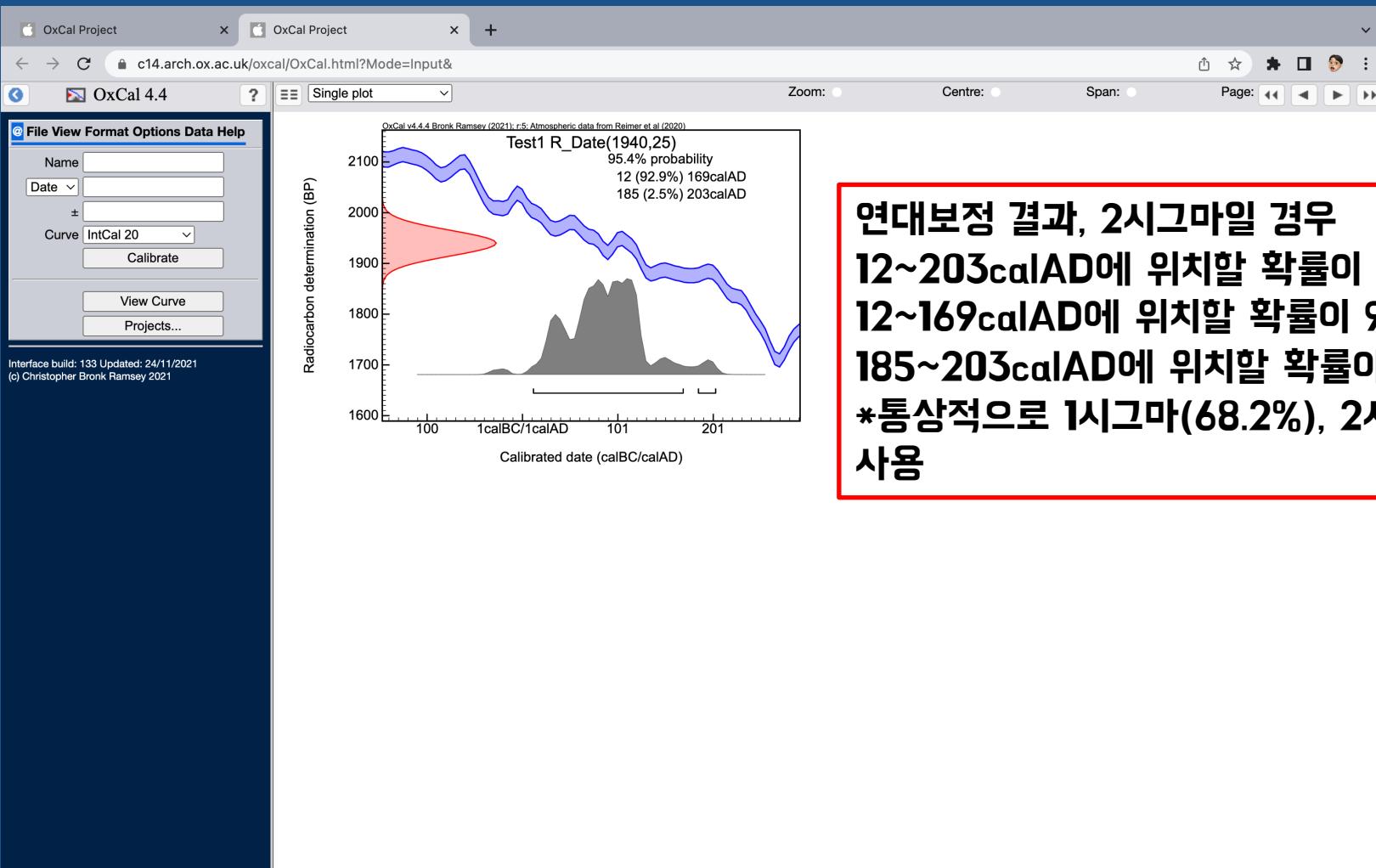
2. 데이터 기입

- Name : 시료명
- 14C Date : 미보정 BP값
- Uncertainty : Error값

*Separator에서 원하는
구분 방식 설정 가능

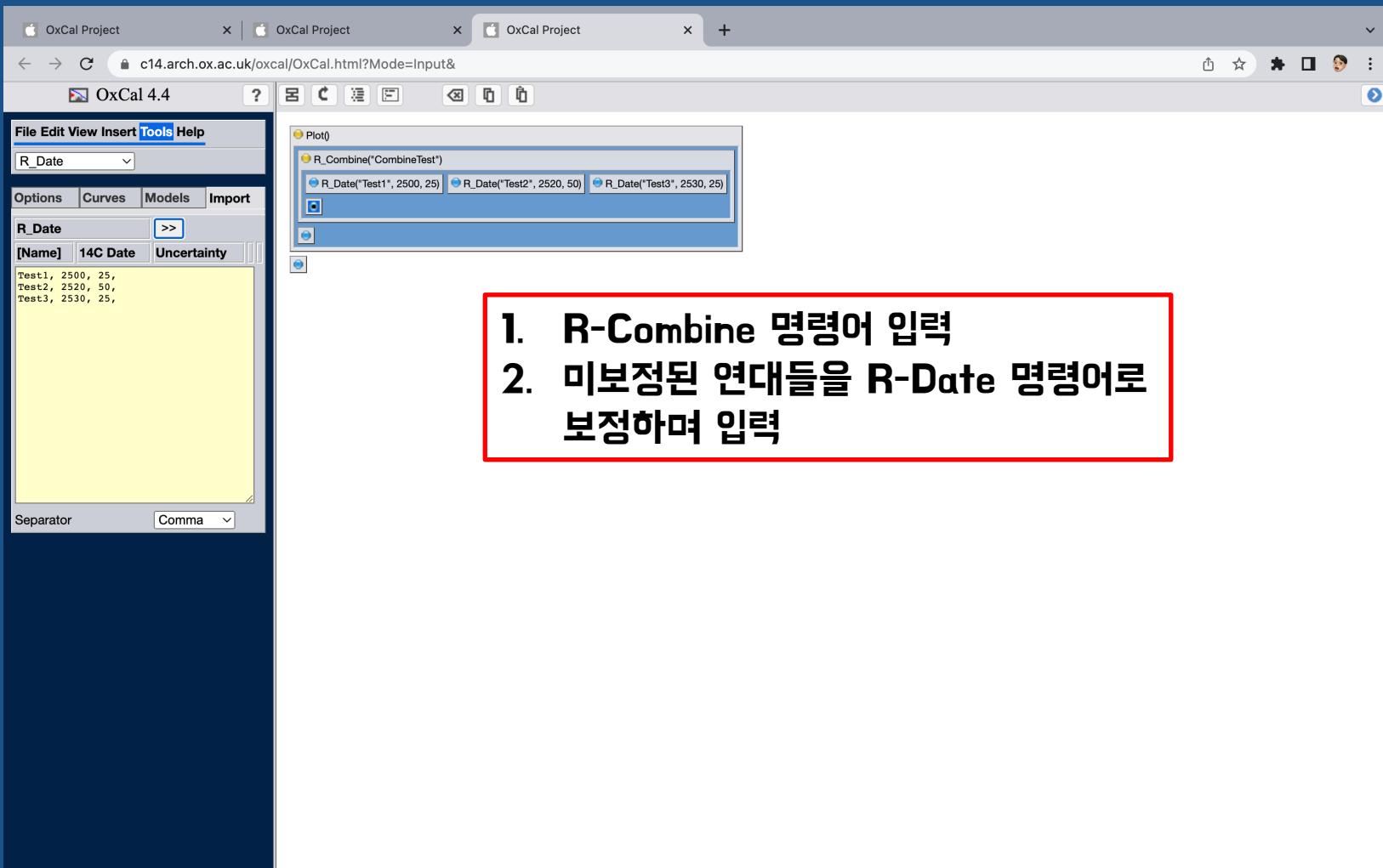


R-Date



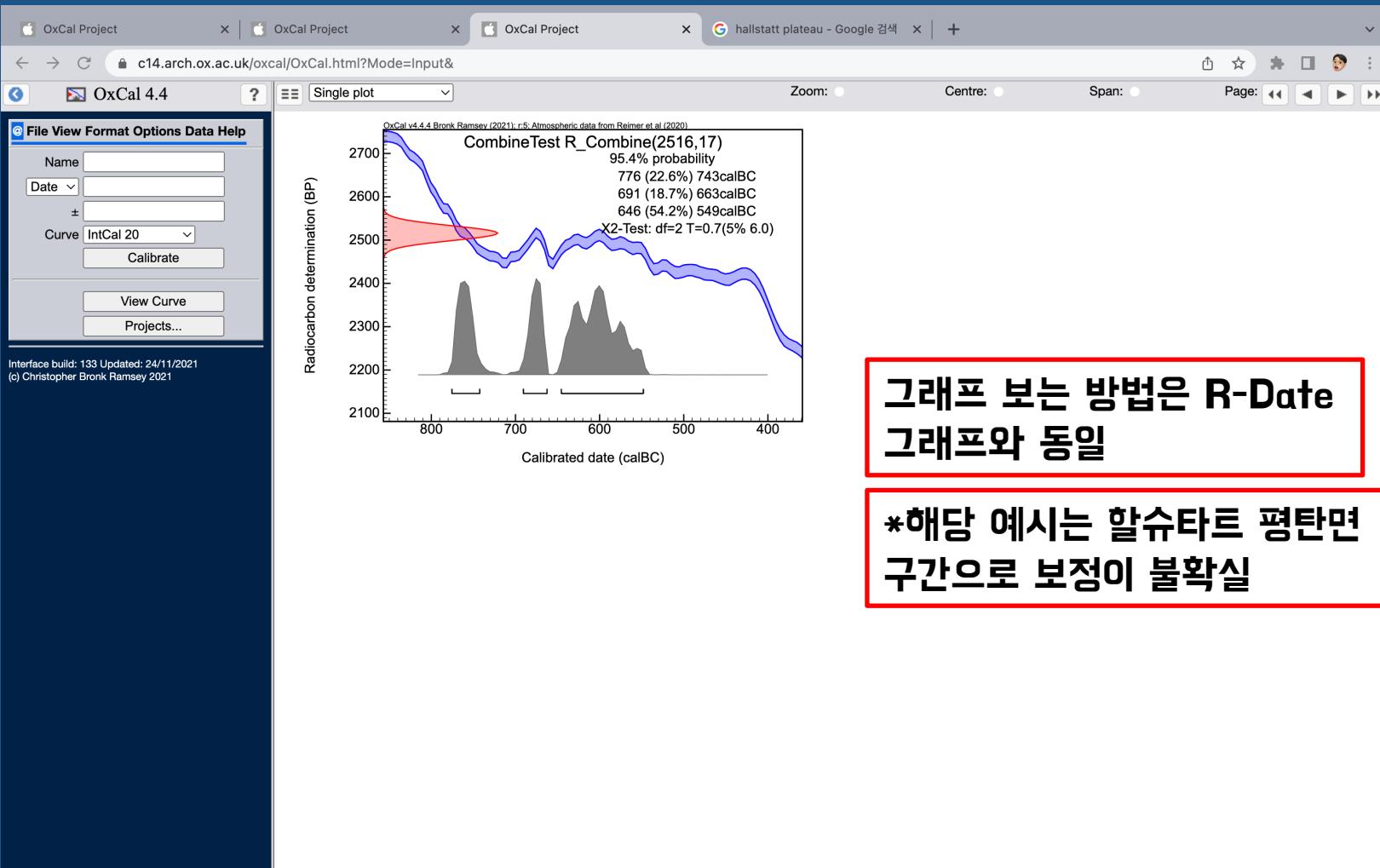
연대보정 결과, 2시그마일 경우
12~203calAD에 위치할 확률이 95.4%
12~169calAD에 위치할 확률이 92.9%
185~203calAD에 위치할 확률이 2.5%
***통상적으로 1시그마(68.2%), 2시그마(95.4%)를 사용**

R-Combine



1. R-Combine 명령어 입력
 2. 미보정된 연대들을 R-Date 명령어로
보정하며 입력

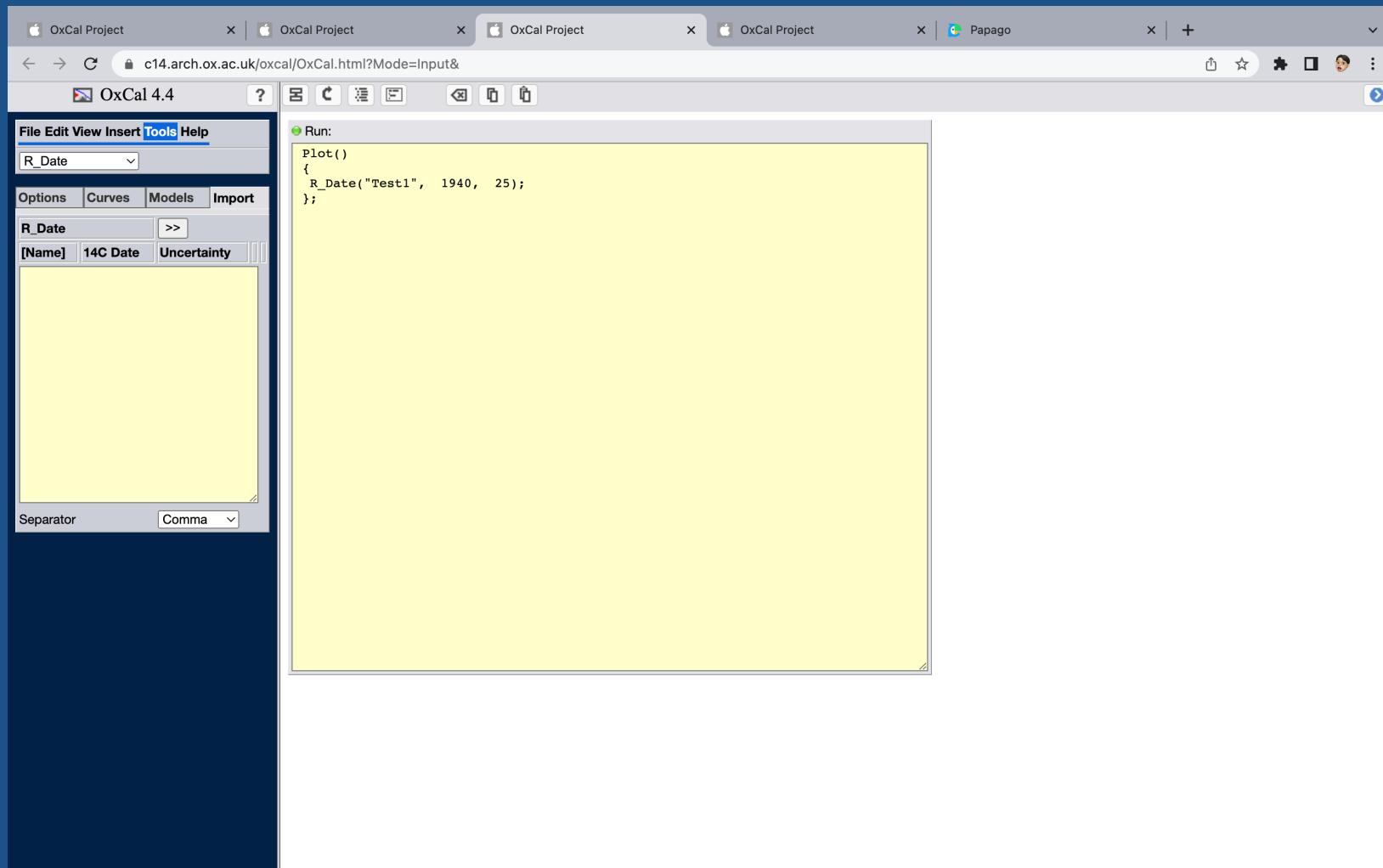
R-Combine



OxCal 스크립트과 OxcAAR

- OxCal은 자체적으로 스크립트를 구성할 수 있는 기능을 제공
- OxcAAR은 R 내에서 OxCAL 스크립트를 실행하고 결과를 불러올 수 있는 기능의 패키지

OxCAL 코딩과 OxcAAR



OxcAAR

```
> quickSetupOxcal()
> 변수명 <- oxcalCalibrate(BP, Error,
  "시료명")
```

*OxcAAR의 간단한 사용법은 아래의 두 링크

참조

https://github.com/ChanToRe/oxcAAR/blob/master/README_ko.md

https://github.com/ChanToRe/oxcAAR/blob/master/vignettes/basic-usage_ko.Rmd

