

Introduction to Statistics

統計学入門

Week 13 | December 22, 2022

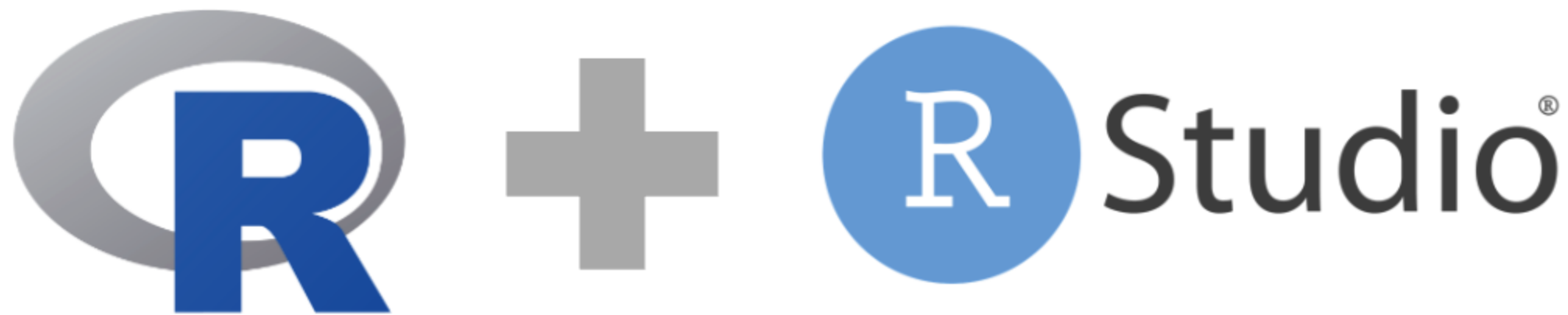
今後のスケジュール

Week	Date	Topic
12	12/15/22（今日）	統計データの入手方法
13	12/22/22	R入門
14	1/12/23	演習（期末テスト対策）
15	1/18/23(水曜11時40分)	期末テスト

始める前に...

Let's take another survey!

今日はR三昧

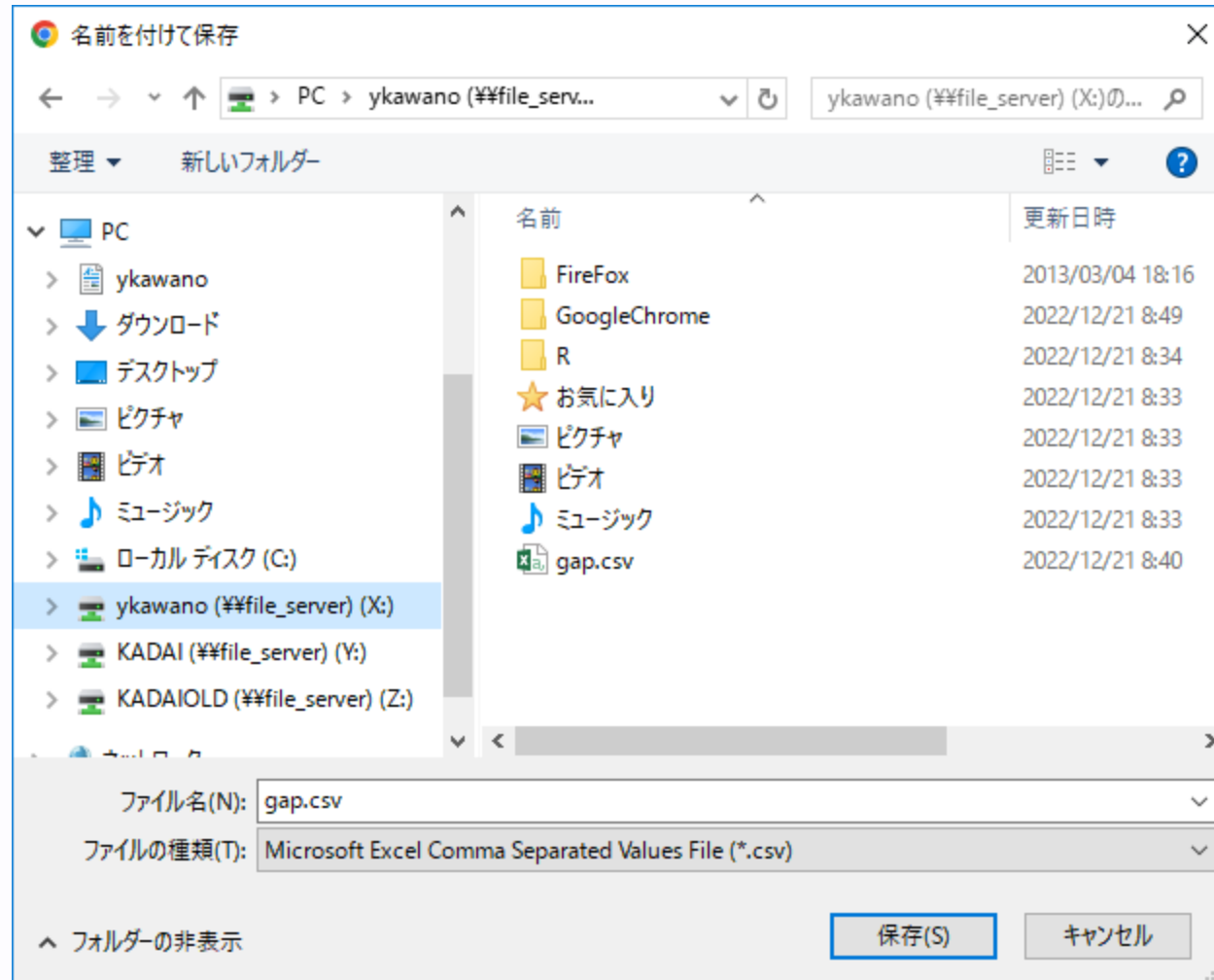




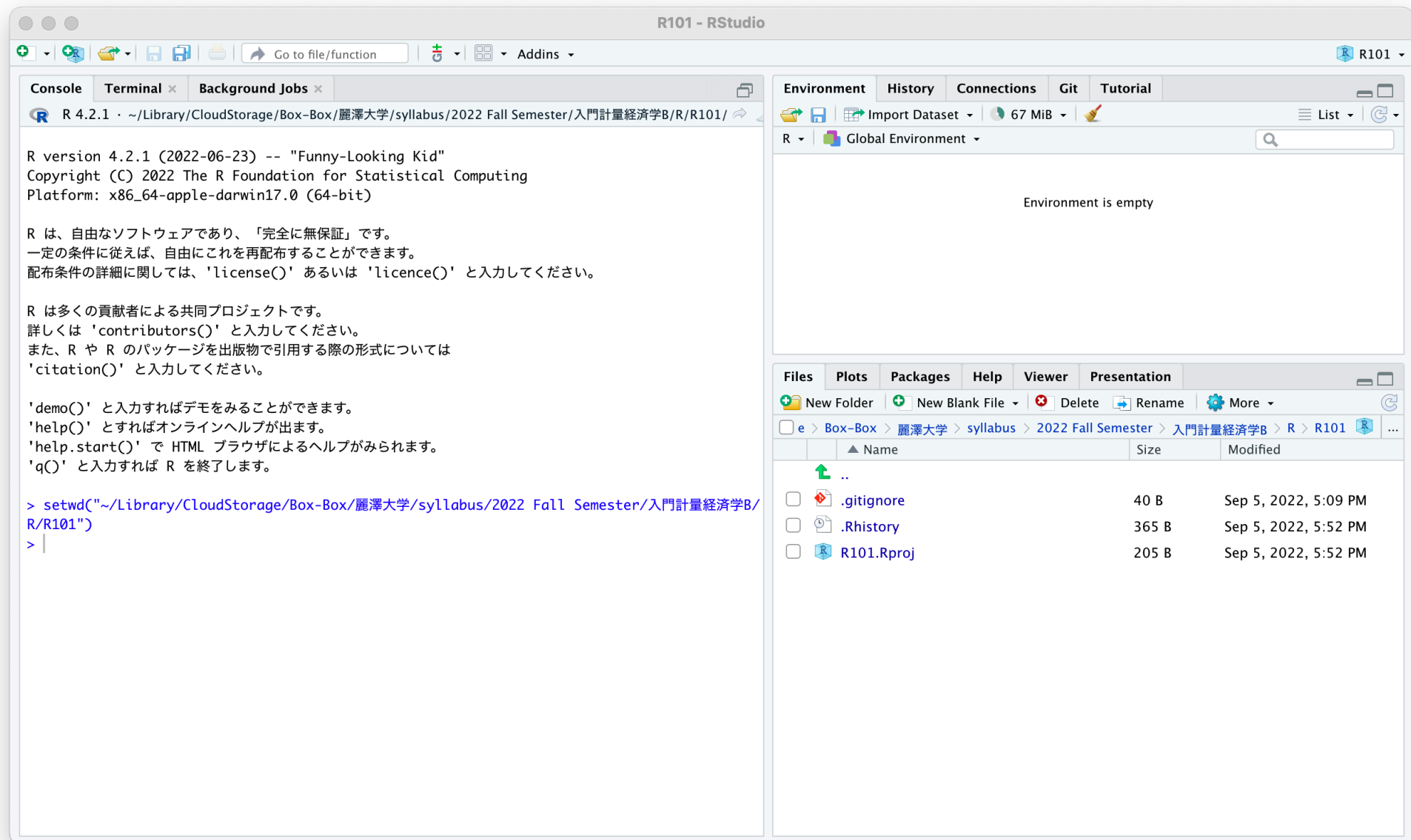
プログラミングの言語

Rプログラミングをやりやすくするためのソフト

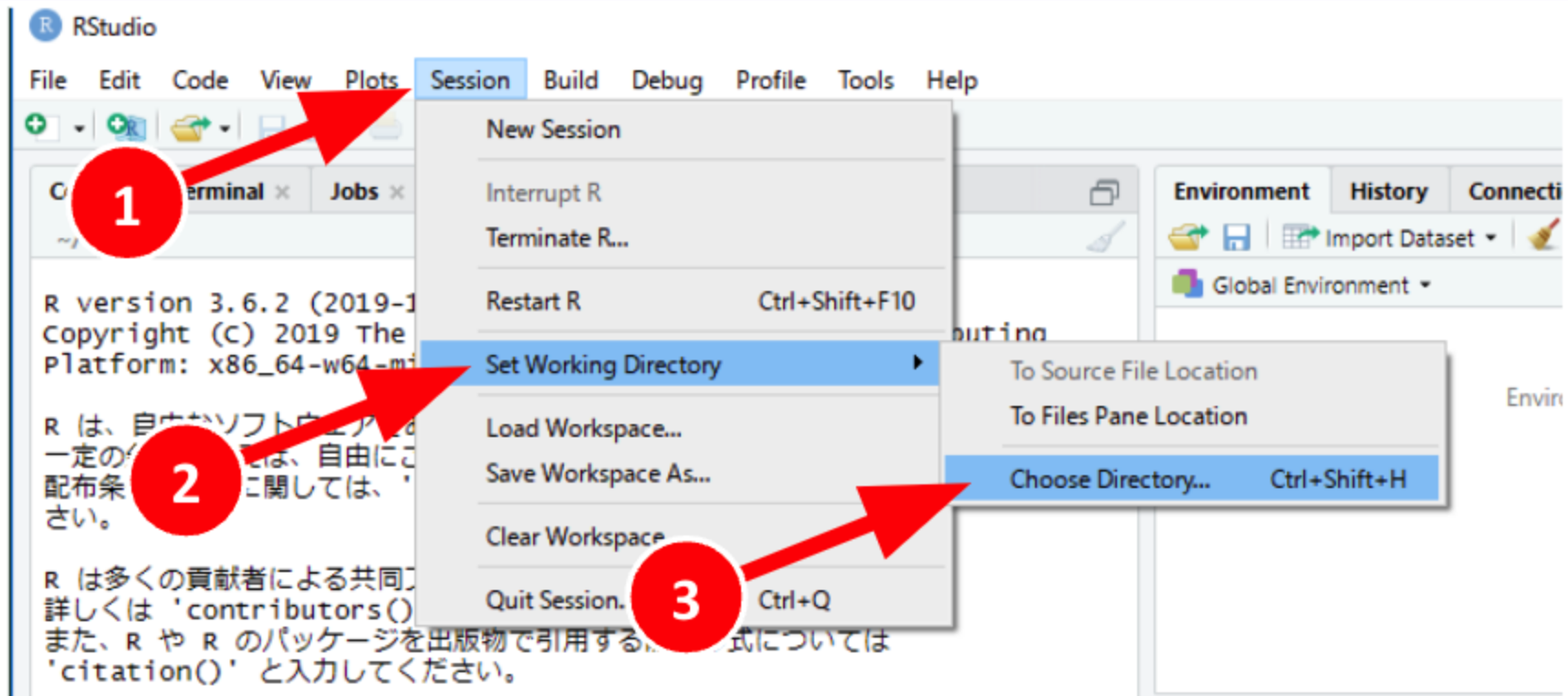
クラスサイトから `class.csv` をX: ドライブにダウンロード



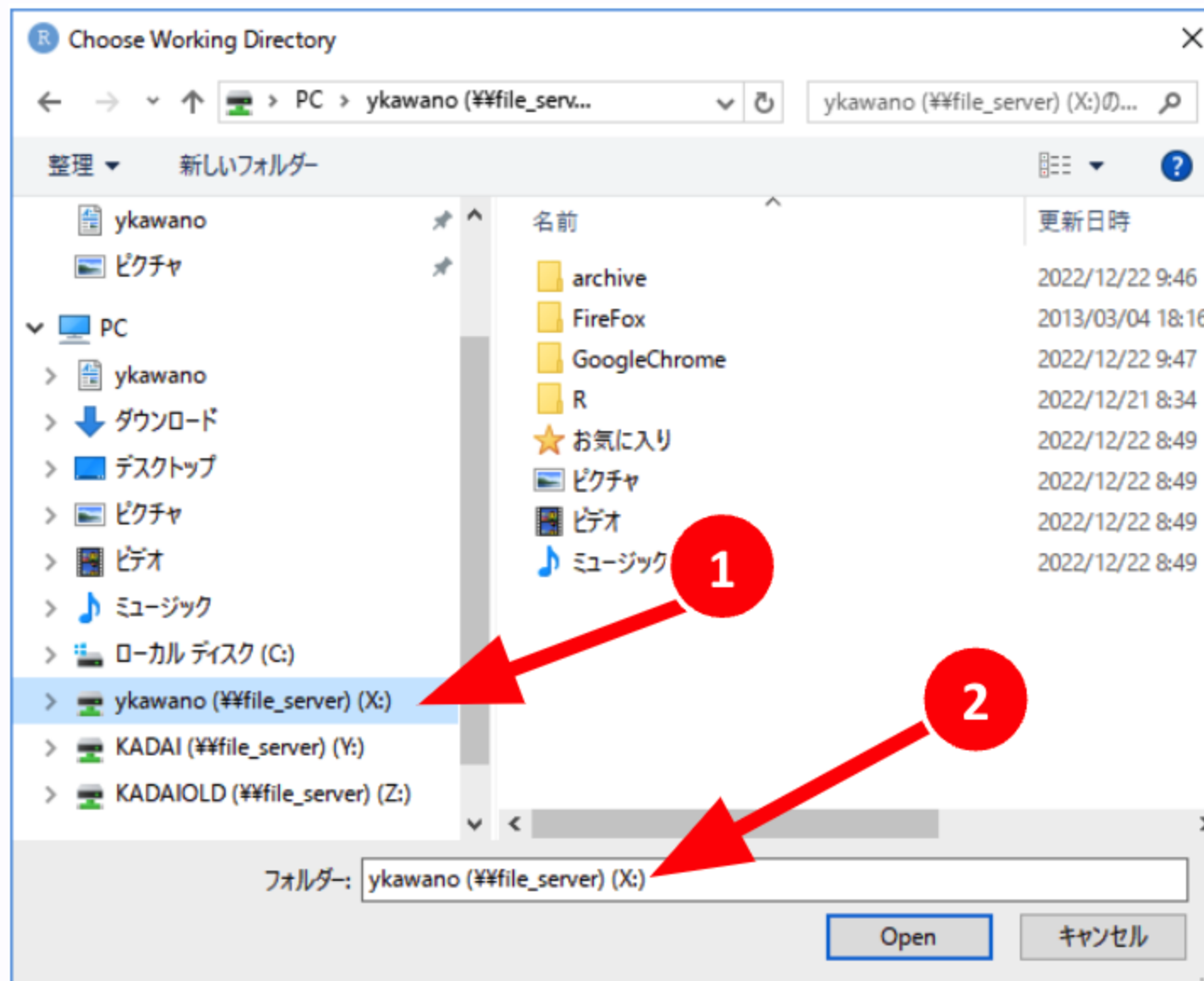
Launch RStudio



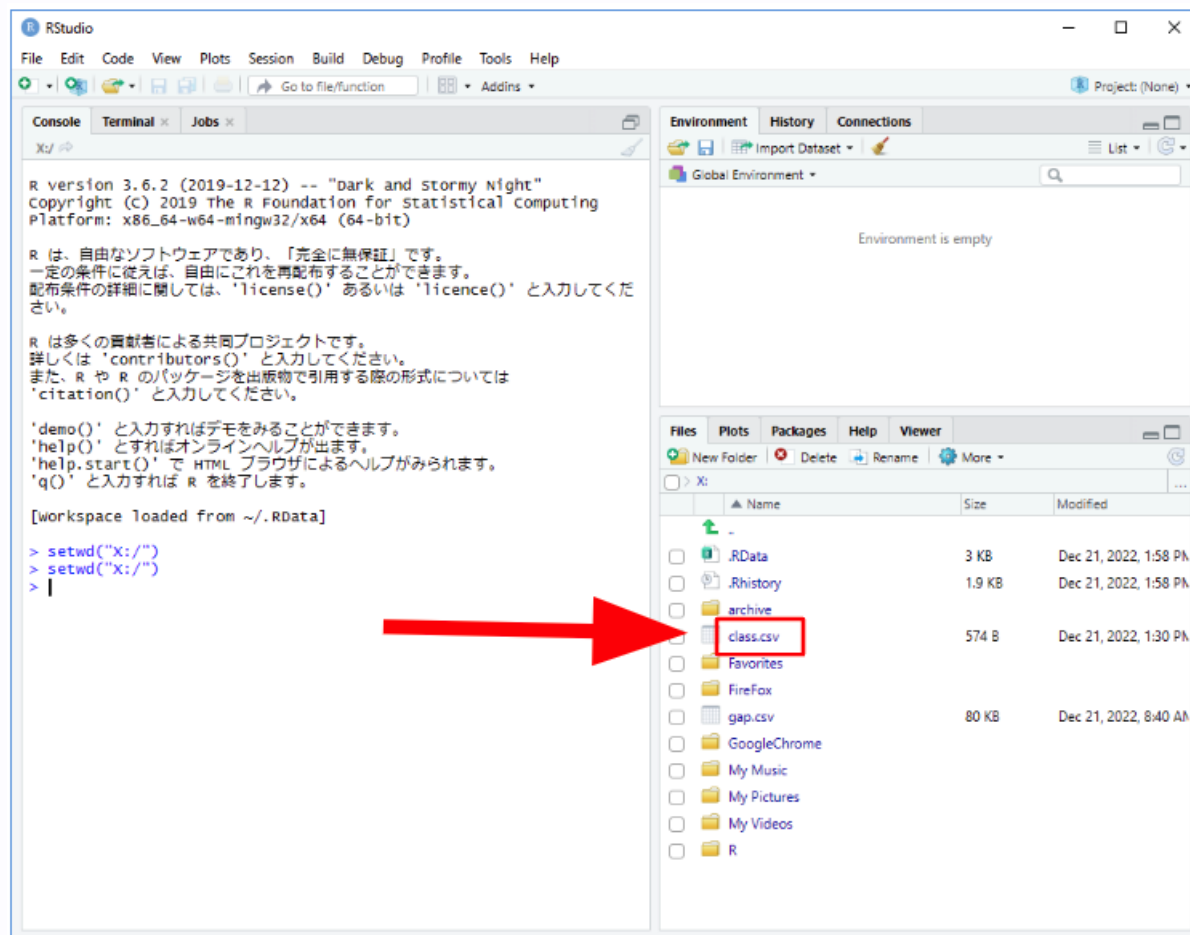
RStudioの「working directory」を設定する



先ほどのX:ドライブを選ぶ



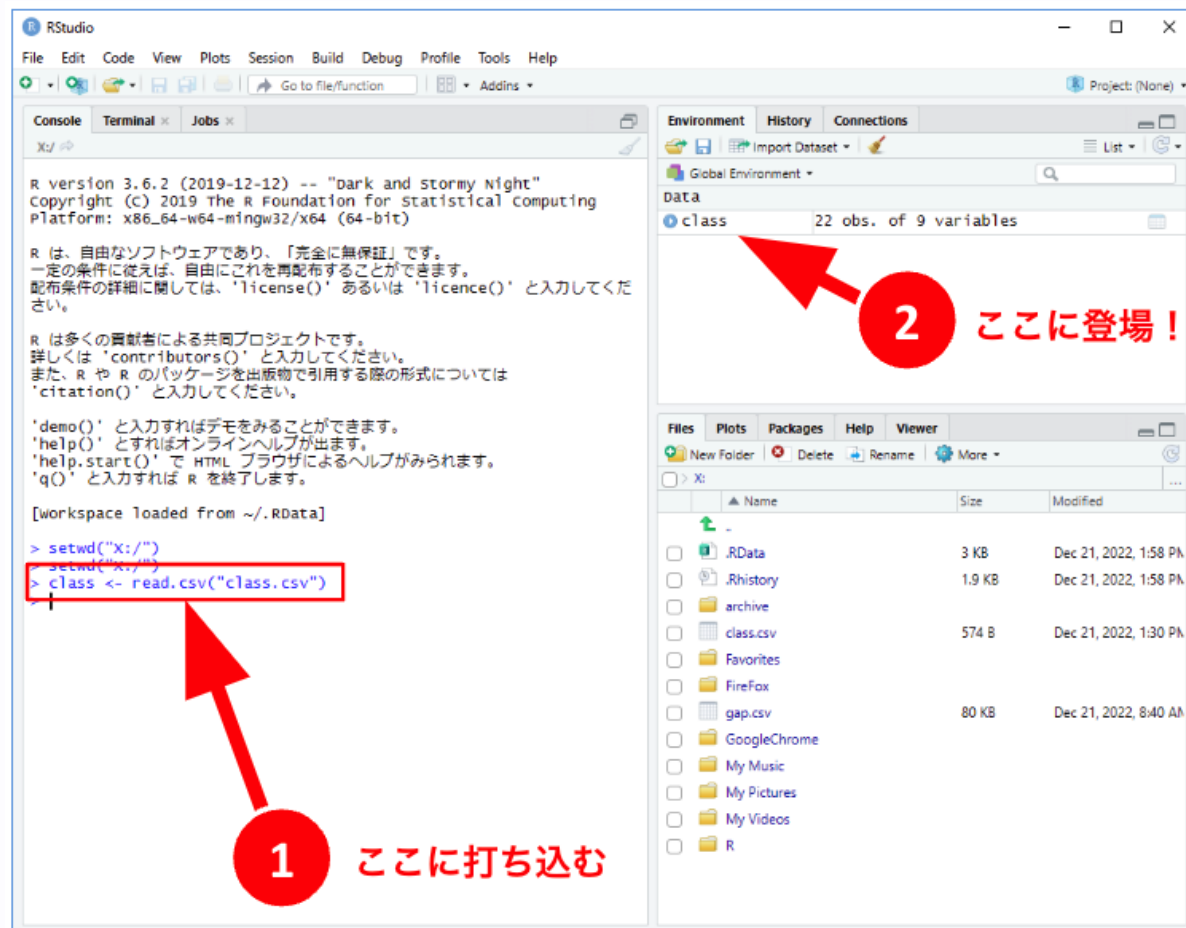
ダウンロードしたclass.csvが入っていることをチェック 



統計タイム！ 🧐
Let's do some stats

データを取り込む

```
class <- read.csv("class.csv")
```



データをインスペクト

次のコマンドを一つずつ実行しよう。

```
class
```

```
summary(class)
```

```
str(class)
```

それぞれ、何が帰ってくるかを把握しましょう！

平均 average

```
mean(class$happiness)
```

`class$` と打ち込むと他の変数が登場！ happiness以外の平均も出してみよう

中央值
median

```
median(class$happiness)
```

標準偏差
standard deviation

```
sd(class$happiness)
```


棒グラフ bar plot

```
barplot(class$happiness)
```

```
barplot(table(class$happiness))
```

この二つの違いは？tableを使うと、プロットはどう変わる？

ヒストグラム histogram

```
hist(class$happiness)
```

箱ヒゲ図

box plot

```
boxplot(class$happiness)
```

グループ分け

```
boxplot(class$happiness~class$partner)
```

色んな組み合わせを試して、それぞれの結果を解釈しよう！

散布図 scatter plot

```
plot(class$happiness~class$study)
```

回帰線 regression line

```
abline(lm(class$happiness~class$study))
```

この二つのコマンドをコンビと考えよう！
他の組み合わせも試そう

ここまでできた人はチャレンジとして、クラスサイトから
`gap.csv`
をダウンロードして同じような分析をしましょう。

gapのデータは年別にフィルターをかける必要がある

```
y2007 <- gap[gap$year==2007,]  
mean(y2007$pop)
```

例えば、上の例だと、2007年のデータを `y2007` 変数に入れてからその平均を出す。

散布図もできます！

ちょっと長いけど、打ち込むといい結果が出るよ！

```
plot(y2007$pop ~ y2007$gdpPercap,  
      col=factor(y2007$continent),  
      pch=19,  
      xlab="GDP per Capita",  
      ylab="Population")
```

大事なのは上の1行目。あとはcolで色別カラムを指定、
pchで色を決め、xlabとylabでラベルをつける

Add a legend

```
legend("topleft",  
      legend=levels(factor(y2007$continent)),  
      pch=19,  
      col=factor(levels(factor(y2007$continent))))
```