

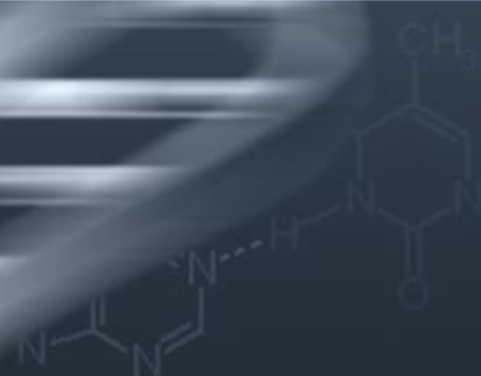
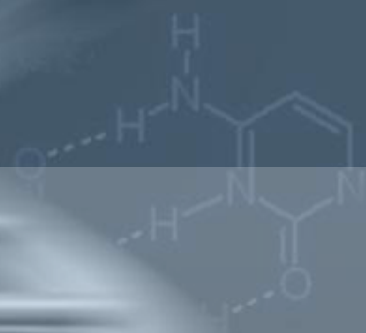


変態統計学入門

～第0回・統計班へようこそ！～

MIS.W 50代 cg/midi/プロ/統計班

shichi



はじめに

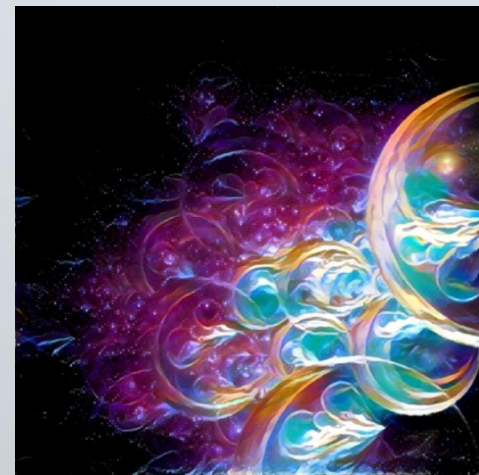
自己紹介



shichi (@shichichaos)

50代。主な活動はCG研である。一年ほど前に前幹事長ことえもえも大先生に肉体改造を施され、気がついたらプログラミングができるようになっていた！！

洗脳前の活動例

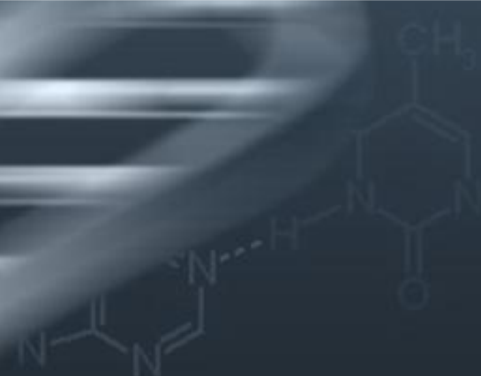
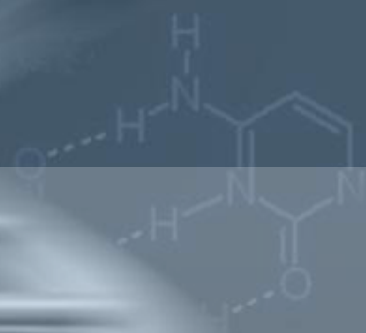


統計班

- ・ **MIS.W**には非公式の班が幾つか存在する
- ・ 現在**50代6人**、**51代5人**で構成
- ・ ガチガチの統計以外もしている
- ・ 卒業研究で使うことになるはず

StockMIS.W

- ・ 株価を予測しよう！って企画
- ・ 実際課金はしない
- ・ 社会人の先輩も在籍 (45代の方も参加してくれました！)
- ・ なんでもできる！ (uiデザイン・webデザイン・解析)



統計学とは？

統計学とは

手持ちのデータ群を分析して意味を持たせる学問

データ



統計的
処理



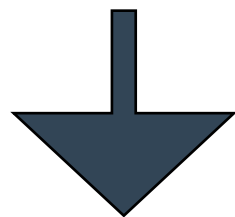
意味

統計学とは

Ex)早稲田生 n 人の属性とGPAのデータ
(属性：オタク・ウェイ・真面目)

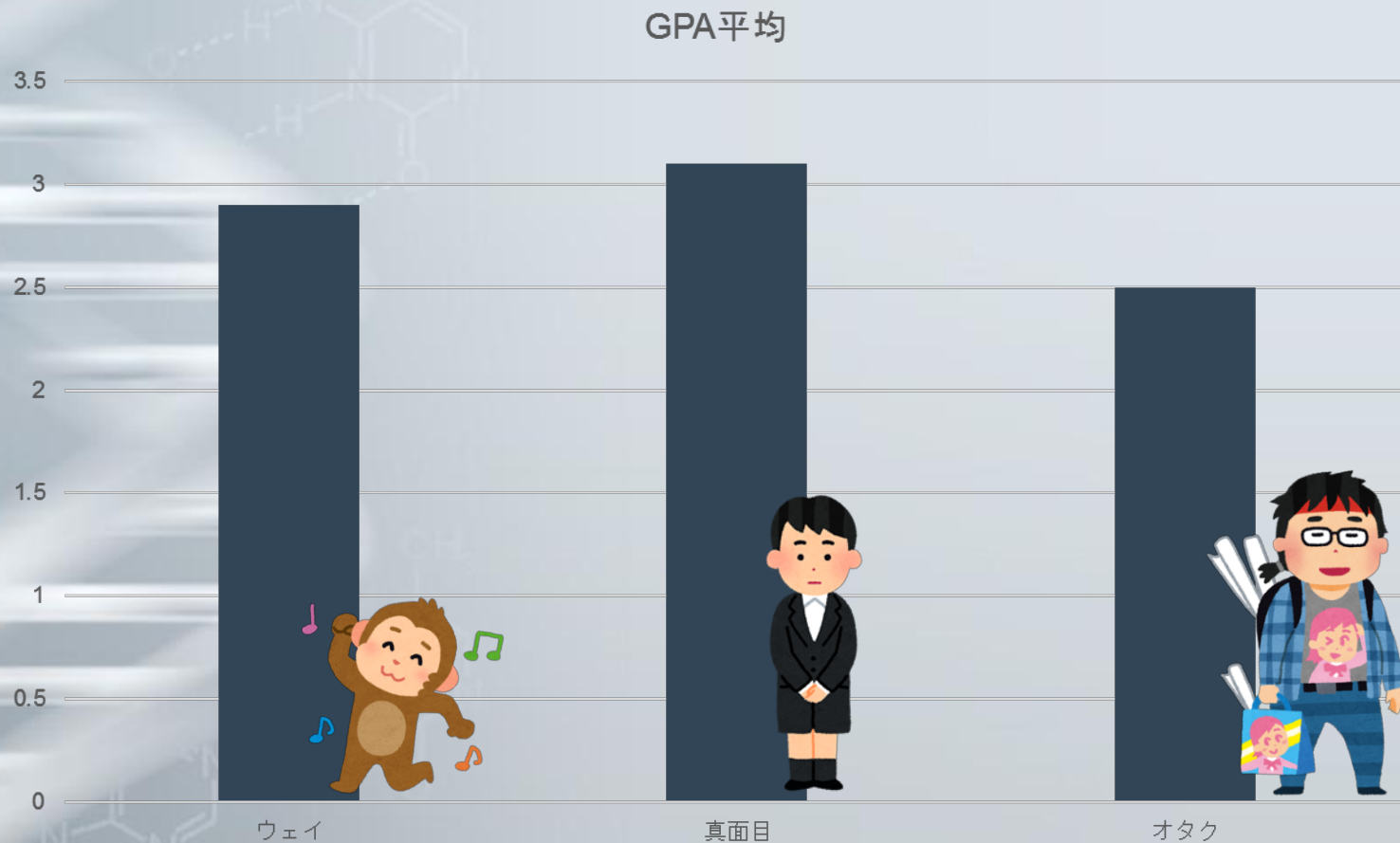


このデータ自体の価値



「へえ～～お前の**GPA**ってそんなくらい
なんだ」程度

ウェイのGPAが意外と高い



統計という「力」



プログラミングという「手段」



大いなる力



R言語とは？

統計解析向けのプログラミング言語

- 統計で使用する様々な処理の多くを数行のコードで実装可能
- **CRAN**アーカイブサイトに使い切れないほどの統計ライブラリ
- 文法が簡単で習得しやすい（言語経験者なら数日でマスター）
- レポートの自動化・**web**公開なども簡単
- **Rstudio**という**IDE**を使うとよりスタイリッシュかつ便利に！

R言語とは

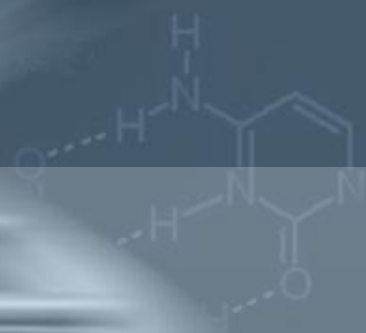
インストール

- R言語

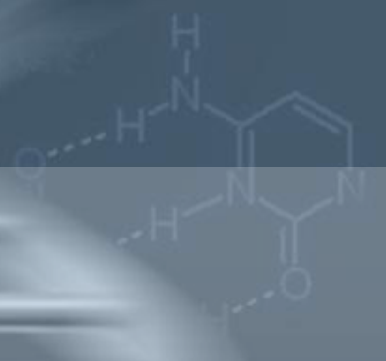
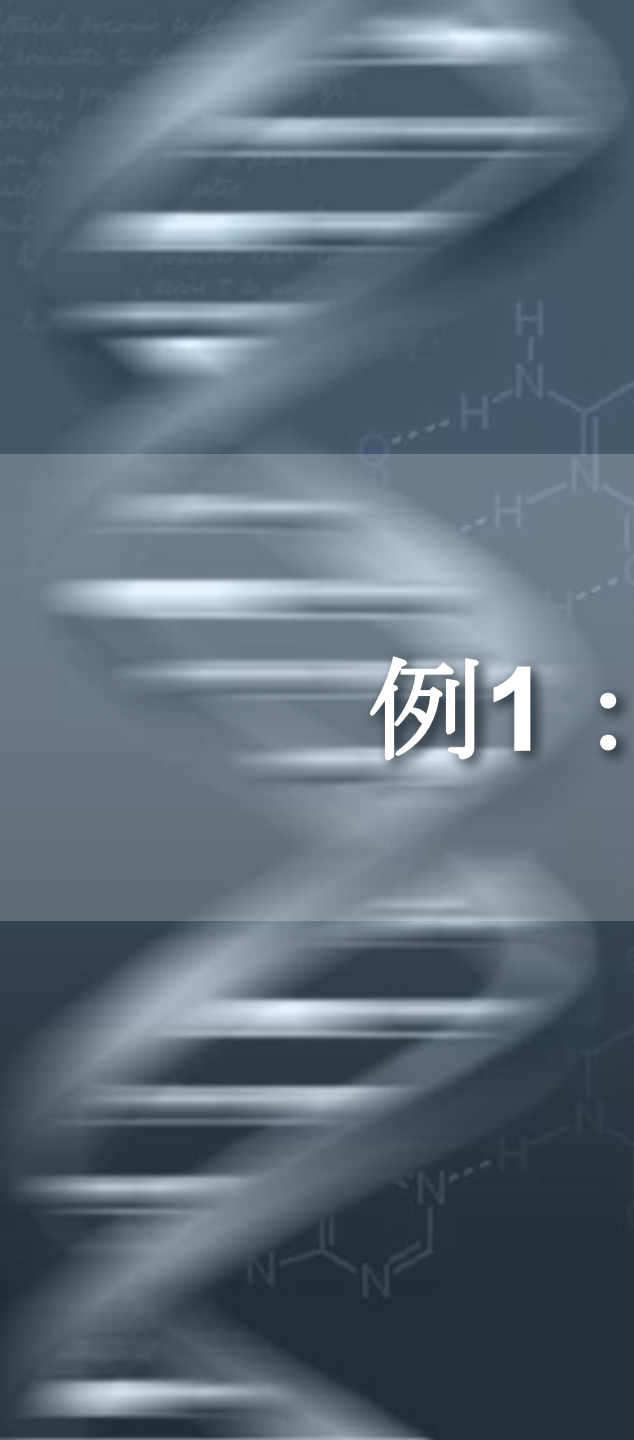
<https://cran.r-project.org/>

- Rstudio

<https://www.rstudio.com/products/rstudio/download/>



活動例



例1：R言語でツイート解析

MIS.Wの人のツイートを解析

ツイート = 言語データ (日本語)



形態素解析

R言語でツイート解析

形態素解析とは？

日本語の文章を、形態素（意味を持つ最小単位）に分割する技術

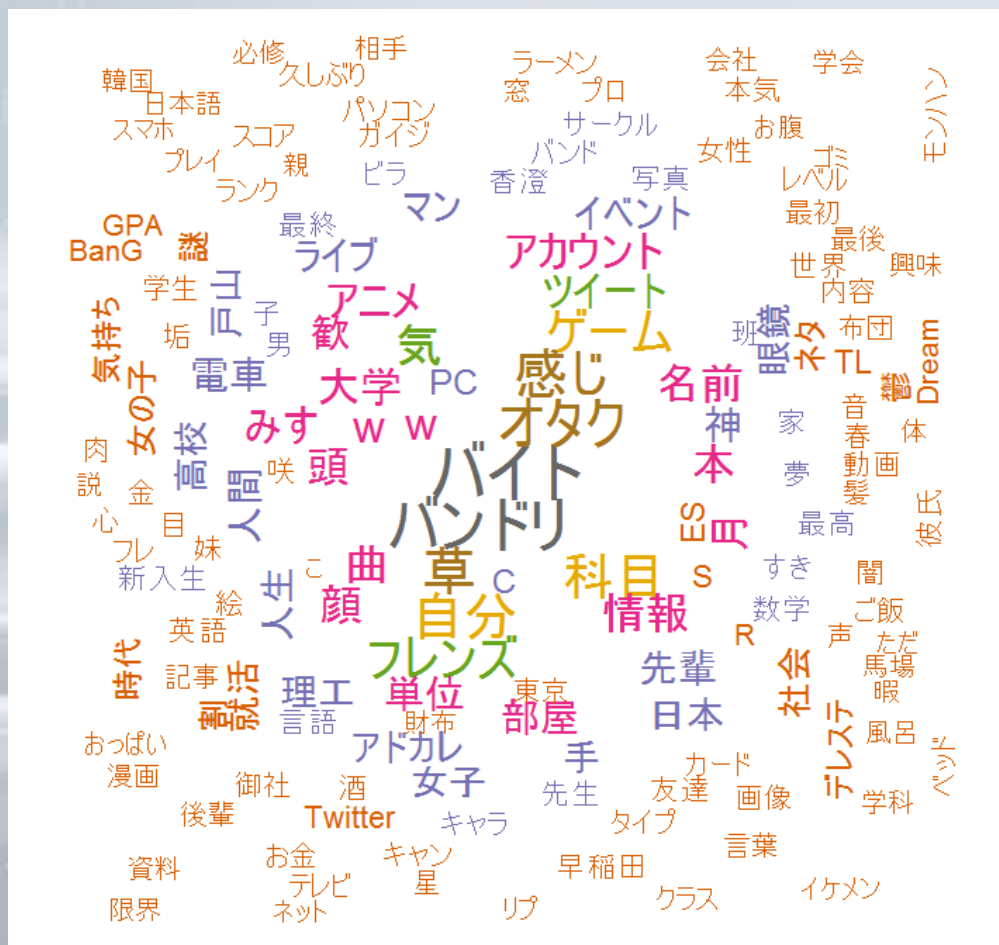
Rならコマンド一発！

(RMeCabを使用)

すっごー
い！

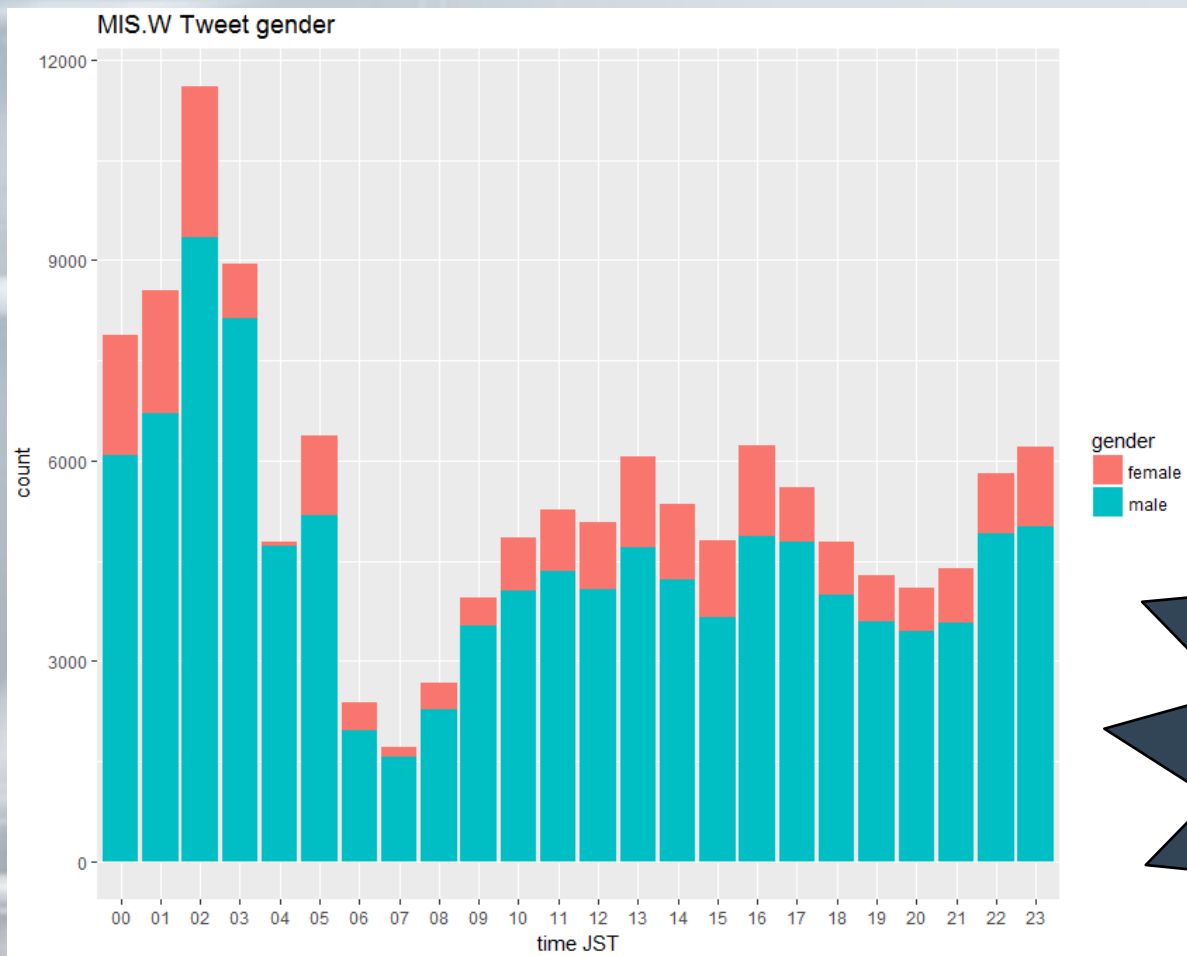
```
> RMeCabC("ようこそ早稲田大学経営情報学会へ！")  
[[1]]  
感動詞  
"ようこそ"  
  
[[2]]  
名詞  
"早稲田大学"  
  
[[3]]  
名詞  
"経営"  
  
[[4]]  
名詞  
"情報"  
  
[[5]]  
名詞  
"学会"  
  
[[6]]  
助詞  
"へ"  
  
[[7]]  
記号  
"!"
```

R言語でツイート解析



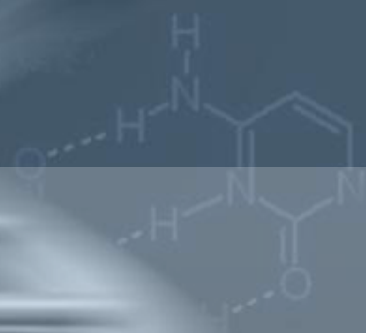
R言語でツイート解析

ggplot2によるツイート頻度ヒストグラム（性別）



非常に高度
な
グラフィッ
ク

詳しくはアドベントカレンダー15日目をご覧ください。

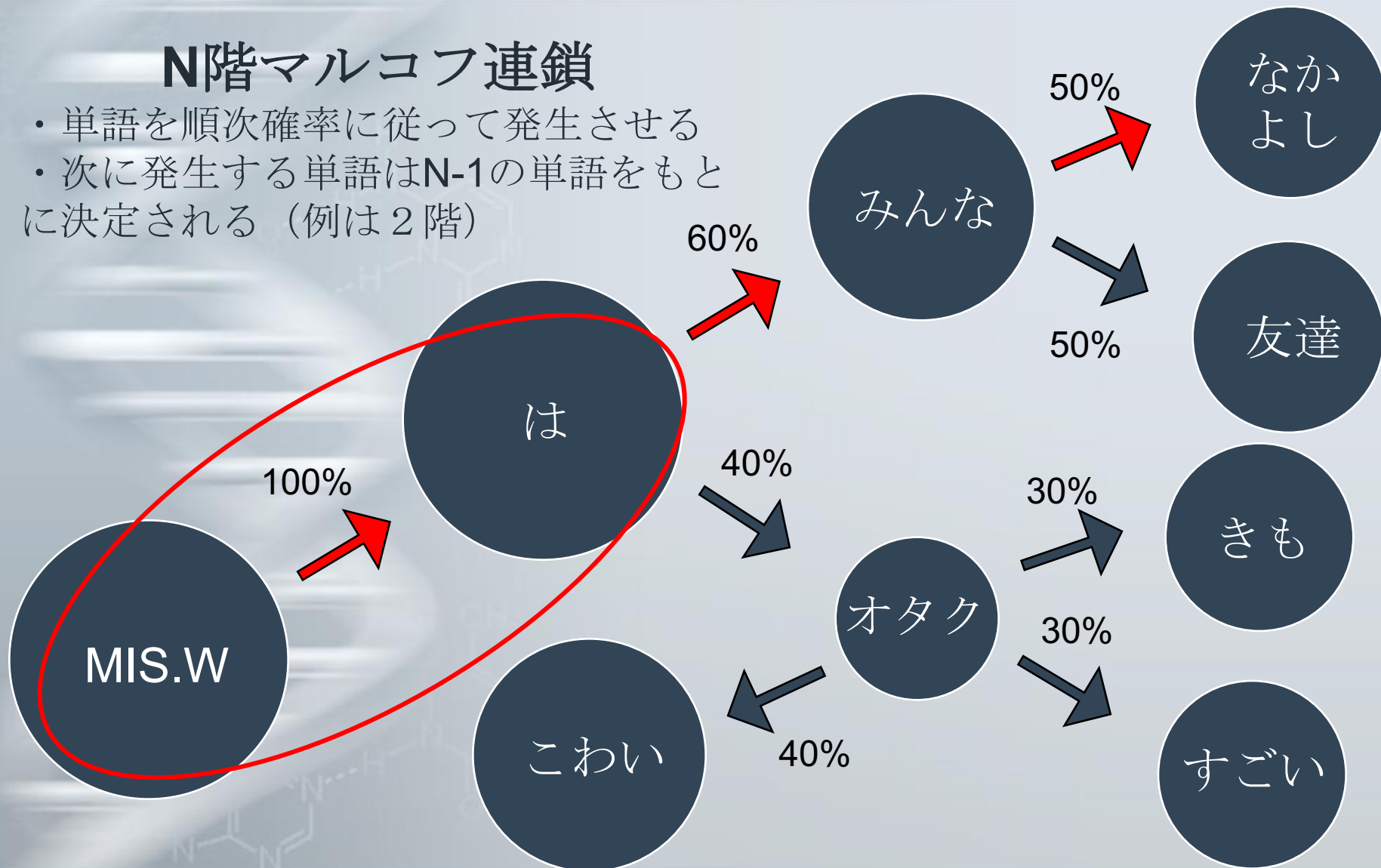


例2：R言語で人工無能bot

R言語で人工無能bot

N階マルコフ連鎖

- ・ 単語を順次確率に従って発生させる
- ・ 次に発生する単語はN-1の単語をもとに決定される（例は2階）



赤い矢印のように単語を生成した結果

「MIS.W は みんな なかよし」

今回の人工無能TweetBotの根本的な部分はこのマルコフ連鎖で作られている。

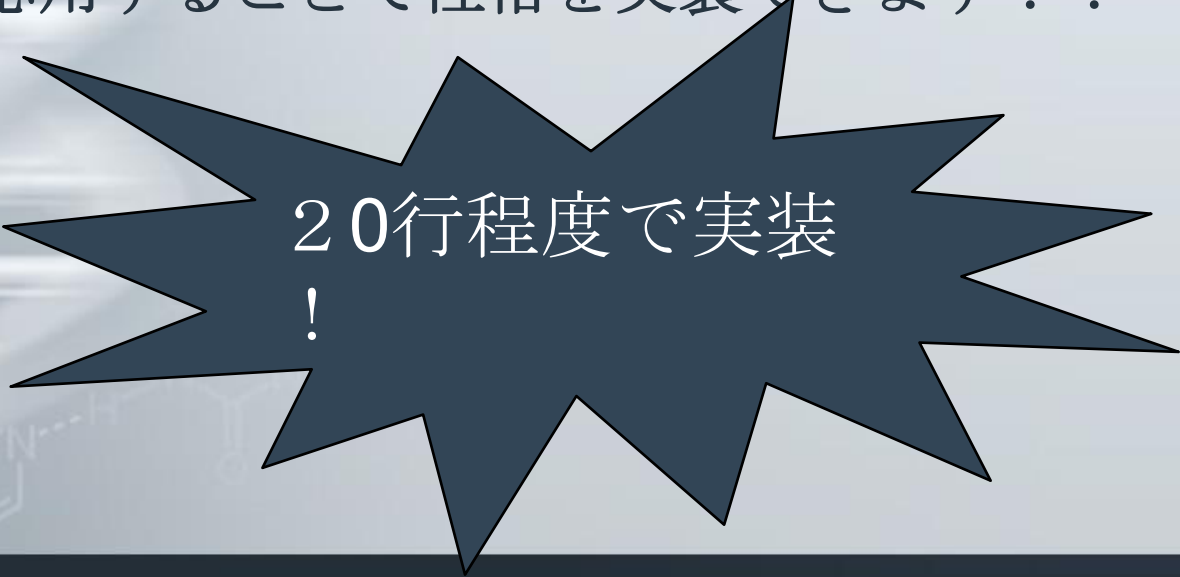
数十行程度で実装！

ナイーブベイズ分類器を用いた性格の実装

ナイーブベイズ分類器の実用例

著者（作曲者）自動判定システム
スパムメール自動判定

応用することで性格を実装できます！！



20行程度で実装
！

R言語で人工無能bot

マルコフ連鎖式
人工無能TweetBot

Spica
@spica_egg



spica @spica_egg · 1月6日

マイコン考察しなくても男です



spica @spica_egg · 1月5日

学会行く度数万飛ぶから、勘弁してほしいようです



2



spica @spica_egg · 1月5日

液卵足りなくて確定反撃(攪拌)出来なくてしんどい



spica @spica_egg · 1月5日

は一あ会社を燃やそう



1



spica @spica_egg · 1月5日

ふ、と気を抜くな、原稿を終えたければな...



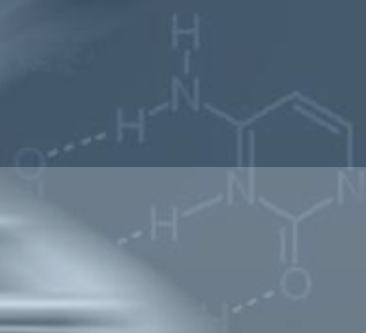
1



spica @spica_egg · 1月5日

嫌がらせコマンドの詰め合わせスクリプト・対象を遥か上空に打ち上げる(地上に落ちるまで10分くらいかかる)





例3：R言語で株価予測

日経225のその日の終値を予測します。

前日の静岡銀行の終値は1024円です。
本日の静岡銀行の株価は1033.07980295566円になるでしょう。
利益率は 0.886699507389155%です。

---,みずほフィナンシャルグループ
前日のみずほフィナンシャルグループの終値は212円です。
本日のみずほフィナンシャルグループの株価は213.105737316264円になるでしょう。
利益率は 0.52157420578473%です。

△,大和証券グループ本社
前日の大和証券グループ本社の終値は725.5円です。
本日の大和証券グループ本社の株価は741.025271012248円になるでしょう。
利益率は 2.13994087005491%です。

---,野村ホールディングス
前日の野村ホールディングスの終値は750円です。
本日の野村ホールディングスの株価は750.300120048019円になるでしょう。
利益率は 0.0400160064025537%です。

---,松井証券
前日の松井証券の終値は937円です。
本日の松井証券の株価は941.017148981779円になるでしょう。
利益率は 0.428724544480171%です。

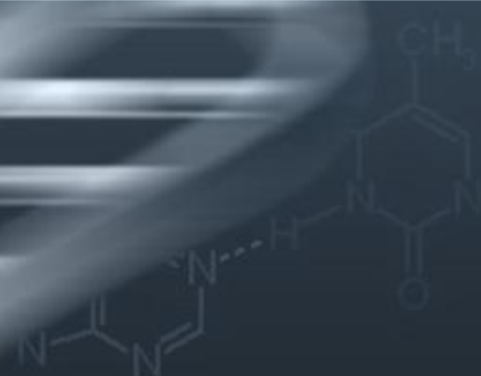
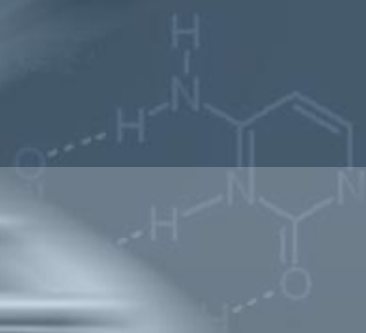
△,損保ジャパン日本興亜ホールディングス
前日の損保ジャパン日本興亜ホールディングスの終値は4350円です。
本日の損保ジャパン日本興亜ホールディングスの株価は4424.24596679916円になるでしょう。
利益率は 1.7068038344634%です。

---,MS&ADインシュアランスグループホール
前日のMS&ADインシュアランスグループホールの終値は3878円です。
本日のMS&ADインシュアランスグループホールの株価は3911.28322496749円になるでしょう。

現状

- ・ 前日比が＋か－かならば7～8割の精度
- ・ しかし幅を考えると**2割**しか...
- ・ 実装は**randomForest**(30行ほど)
- ・ **web**上の記事を解析して補正する試み（実装中）

→求ム班員・企画員！！



統計班への入り方

統計班への入り方

- slackのmis_stat窓にjoinする
- 統計班の班員または班長(@shichichaos)に入会の意を示す
- MisWikiのプロフィールに統計班とタグ付けする

のいずれか

次回は統計基礎&R講座の「変態統計学入門①」でお会いしましょう。