



情報工学実験Ⅱ

第6章 スクリプト言語を用いた テキストファイル処理に関する実験

担当：岩田



■ ■ ■ 本実験の目的

- スクリプト言語を使ったプログラミングを学ぶ
 - Python(パイソン)を使う
- できるようになること
 - データ整理
 - データ加工
 - テキストファイル処理



なぜPythonなのか？

➤ スクリプト言語(インタプリタ)ならほかにもいっぱいあるじゃないか！ そっちを使わせろ！

□ バッチファイル/シェルスクリプト

□ Perl

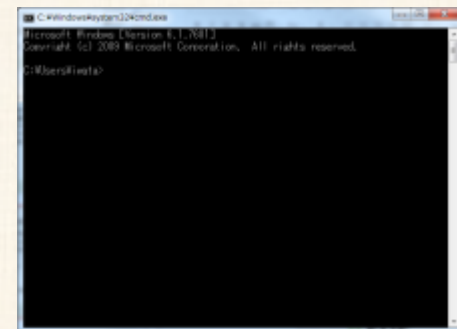
□ Ruby

□ PHP

□ Lua

□ AWK/sed/grep

□ VBS



など



なぜPythonなのか？

➤ 読みやすくて書きやすい！

□ 制御構造をインデントで表現する（ちゃんとやろう）

```
def factorial(x):  
    if x == 0:  
        return 1  
    else:  
        return x * factorial(x - 1)
```

インデントでブロックを表現

pythonの例

```
public static int factorial(int x)  
{  
    if (x == 0) {  
        return 1;  
    } else {  
        return x * factorial(x - 1);  
    }  
}
```

中括弧でブロックを表現

分かりやすいJavaの例

なぜPythonなのか？

➤ 機械学習のデファクトスタンダード

□ 参考

- ✓ Python以外でもできる。言語別ディープラーニング用フレームワークまとめ

- ◆ <https://blog.tkrel.com/12579>

- JavaScript, Swift, Ruby, Perl, Java, PHPなど

➤ さまざまな利用例

参考:

<https://www.python.org/about/quotes/>

- Google

- YouTube.com

スクリプト言語の使いどころ (1/2)

➤ 書き捨てるコード

□ たとえば！

✓ 「情報の学生の就職先ってどこが一番多いの？」を解決

情報工学科 卒業生就職先一覧

(第1期生 ~ 平成27年3月末現在) <https://www.info.nara-k.ac.jp/carrier/corporation.html>

| 番号 | 企業名 | 卒業年度 | | | | | | | | | | | | | | | | | | | | | | | | |
|----|-------------|------|----|----|----|----|----|----|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| | | H2 | H3 | H4 | H5 | H6 | H7 | H8 | H9 | H10 | H11 | H12 | H13 | H14 | H15 | H16 | H17 | H18 | H19 | H20 | H21 | H22 | H23 | H24 | H25 | H26 |
| 1 | ローム株式会社 | 1 | 1 | 1 | | | | | | | | | | | | | | | | | | | | | | 合計3 |
| 2 | 日本電気株式会社 | 1 | | | | | | | | | | | | | | | | | | | | | | | | 合計1 |
| | 富士通システム株式会社 | | | | | | | | | | | | | | | | | | | | | | | | | |

面倒すぎる！

よしたに著 “理系の人々”

http://blogs.yahoo.co.jp/engineer_ryuseigun/11418770.html

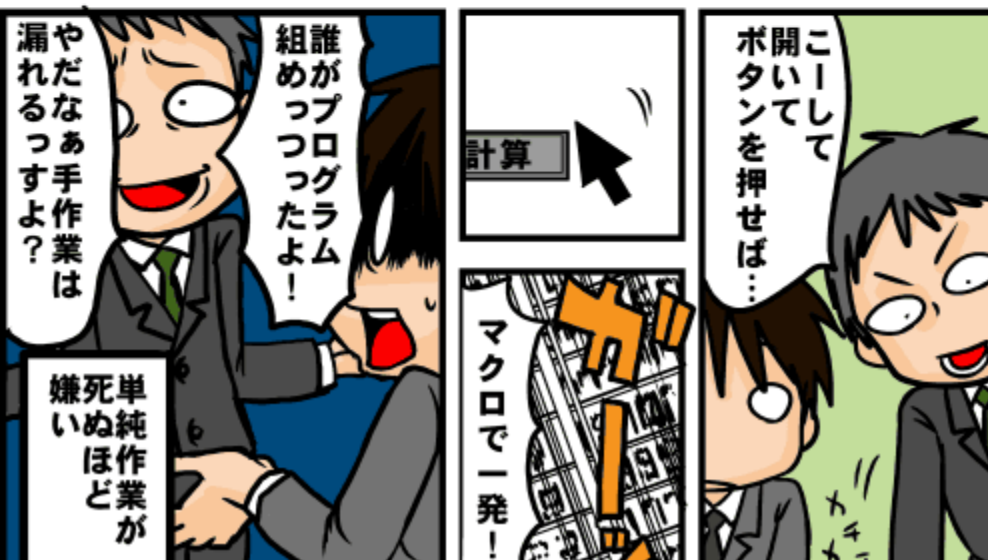
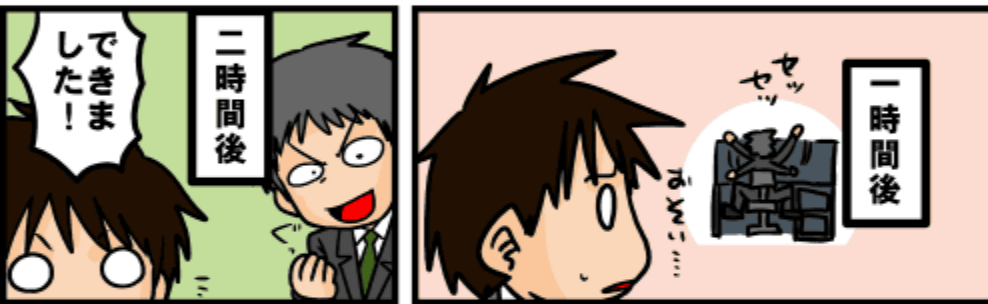
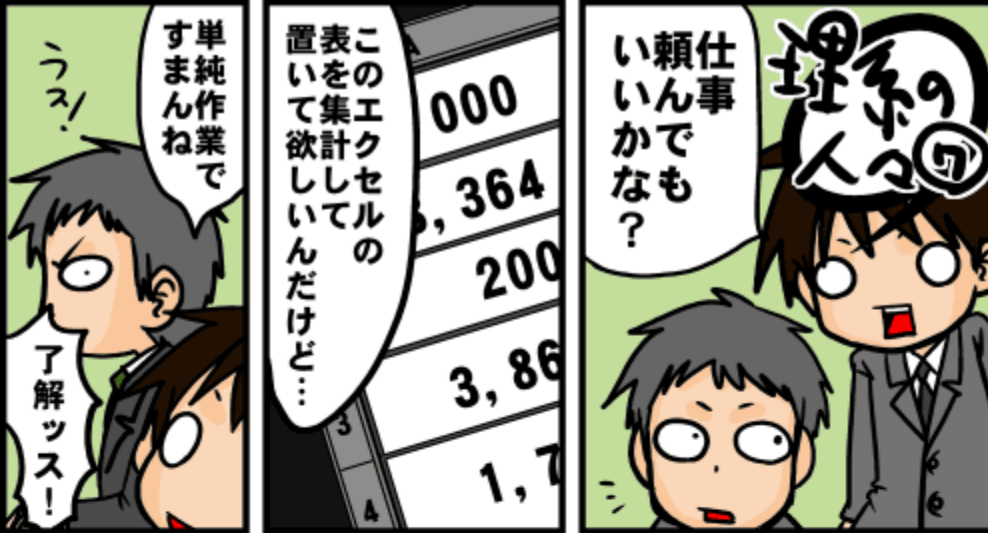
➤ 単純作業のデメリット

□ 面倒

- ✓ インテリジェンスを感じない
- ✓ PCができる仕事は任せる
- ✓ もう一回この作業しなきゃならないとか耐えれない

□ 漏れ落ちる

- ✓ 人間は必ずミスる
- ✓ ちゃんと書けば機械はミスらない



■ スクリプト言語の使いどころ (2/2)

➤ 実験の自動化

- そこそこ時間がかかる実験を連続してやる必要がある

- ✓ 処理A: 3h

- ✓ 処理B: 4h

- ✓ 処理C: 1h

※複数ファイルの動画のエンコードとか
RAW画像の現像とか

- 待ってられるか！

- ✓ 処理Aが終わったら処理Bを始めて、終わったら処理Cを始めるプログラムを書けば、寝てる間(8h)に実験が終わる！

- バッチファイルなんて言ったりする

さっそく使ってみよう(自分のPCの場合)

➤ まずはPythonをインストール

python.org へ移動 (検索サイトからでもOK)

python™

Download for Windows

Python 3.8.3

Note that Python 3.5+ cannot be used on Windows XP or earlier.

DownloadsからPython 3.8.3をダウンロード

Python is a programming language that lets you work quickly and integrate systems more effectively. >>> [Learn More](#)

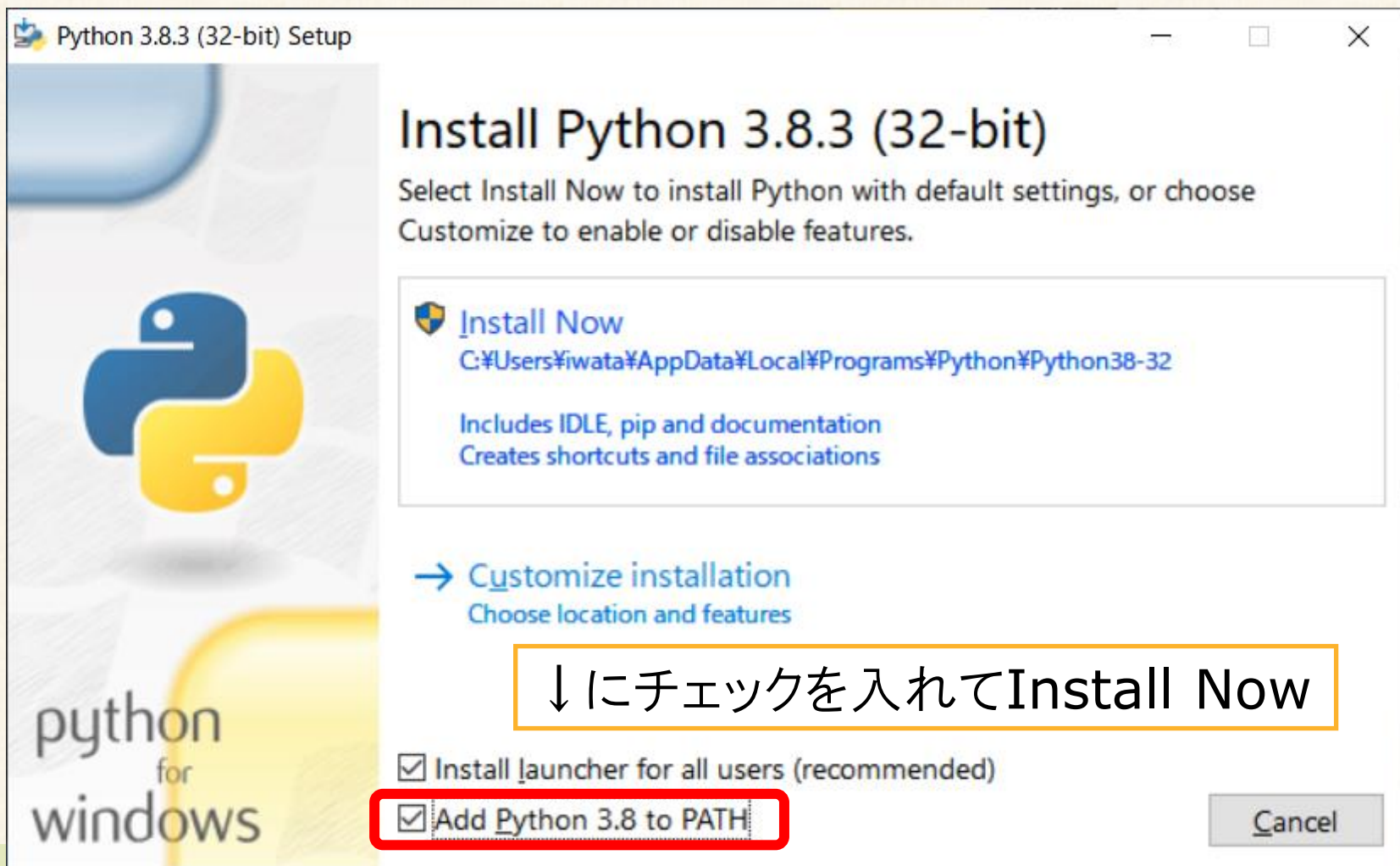
ダウンロードしたファイルをインストール

この種類のファイルはコンピュータに損害を与える可能性があります。python-3.8.3.exe のダウンロードを続けますか?

保存 破棄

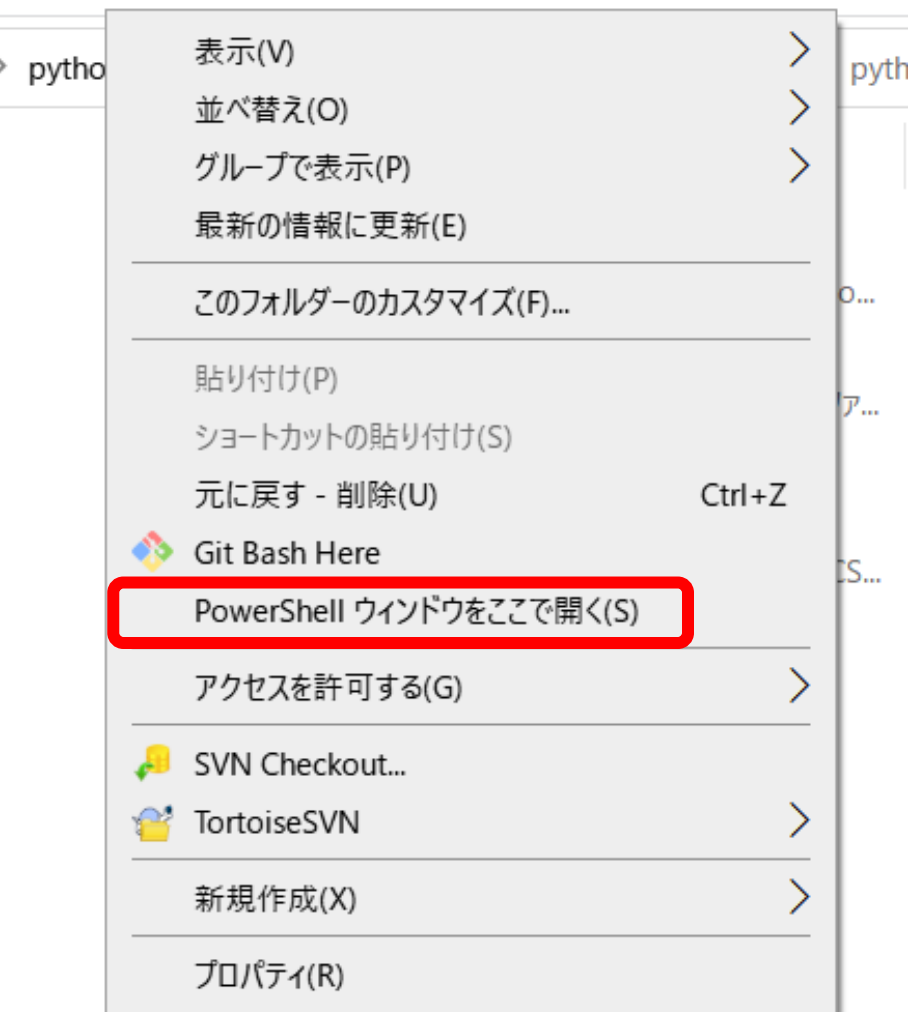
すべて表示

Pythonのインストール(自分のPCの場合)



インストールが終わったらCloseを押す (PATH LimitはどちらでもOK)

Pythonを実行してみよう



1. 「ドキュメント」フォルダにPythonフォルダを作る
(別のフォルダでもOK)
2. Teamsのファイルタブからsample.pyをコピーしてくる
3. Pythonフォルダで、Shiftキーを押しながら右クリックすると、「PowerShellウィンドウをここで開く」が出てくる

Hello World

> **python sample.py** をターミナルで実行する

※途中まで打ってTabキーで補間できるので便利に使いましょう
上キーを押したら前のコマンドが出てくるので、再実行時に便利です

```
PS C:\Users\iwata\Documents\svn\2020\exp2\python-sample> dir
```

```
ディレクトリ: C:\Users\iwata\Documents\svn\2020\exp2\python-sample
```

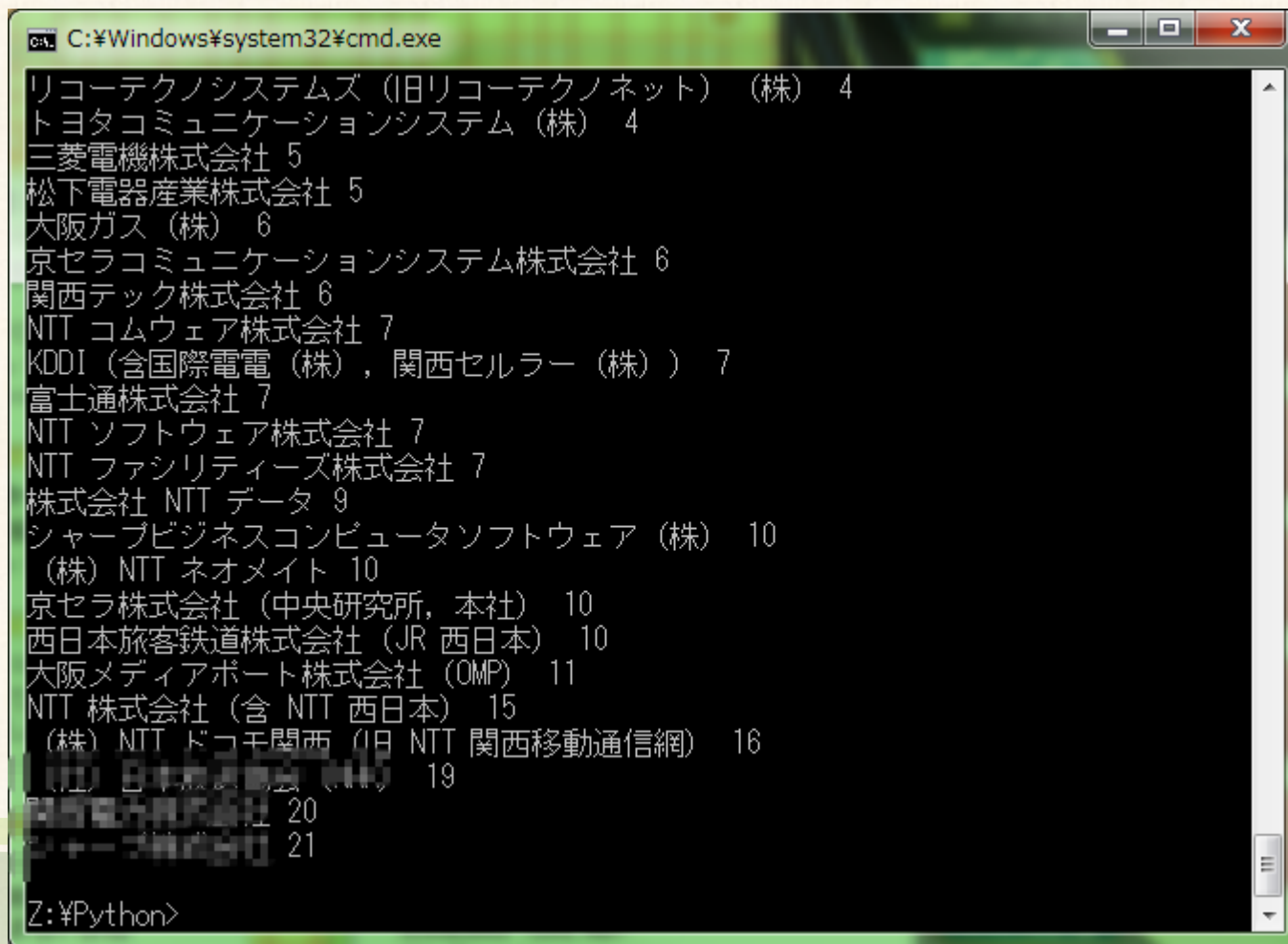
| Mode | LastWriteTime | Length | Name |
|--------|------------------|---------|--------------------|
| -a---- | 2019/02/08 18:56 | 1326 | corp.py |
| -a---- | 2019/02/08 18:56 | 278222 | corporation.html |
| -a---- | 2019/02/08 18:56 | 384 | k031.py |
| -a---- | 2019/02/08 18:56 | 389 | mp42carnavi v2.bat |
| -a---- | 2019/12/19 23:04 | 3682008 | overlap.dat |
| -a---- | 2019/02/08 18:56 | 103 | sample.py |
| -a---- | 2019/02/08 18:56 | 2806 | title_list.csv |

```
PS C:\Users\iwata\Documents\svn\2020\exp2\python-sample> python .\sample.py
Hello World
```

```
PS C:\Users\iwata\Documents\svn\2020\exp2\python-sample>
```

一番多い就職先はどこですか？

➤ corp.pyを実行して確かめてみてください



```
C:\Windows\system32\cmd.exe
リコーテクノシステムズ (旧リコーテクノネット) (株) 4
トヨタコミュニケーションシステム (株) 4
三菱電機株式会社 5
松下電器産業株式会社 5
大阪ガス (株) 6
京セラコミュニケーションシステム株式会社 6
関西テック株式会社 6
NTT コムウェア株式会社 7
KDDI (含国際電電 (株), 関西セルラー (株)) 7
富士通株式会社 7
NTT ソフトウェア株式会社 7
NTT ファシリティーズ株式会社 7
株式会社 NTT データ 9
シャープビジネスコンピュータソフトウェア (株) 10
(株) NTT ネオメイト 10
京セラ株式会社 (中央研究所, 本社) 10
西日本旅客鉄道株式会社 (JR 西日本) 10
大阪メディアポート株式会社 (OMP) 11
NTT 株式会社 (含 NTT 西日本) 15
(株) NTT ドコモ関西 (旧 NTT 関西移動通信網) 16
(株) 日本放送協会 (NHK) 19
関西電力株式会社 20
シャープ株式会社 21
Z:\Python>
```

■ チェックポイントについて(40点分)

- 内容を理解しているか教員が確認する
 - 作成したソースコードと実行結果, アルゴリズムなどについて説明してください
 - ✓ 確認結果は学生の皆さんも[こちらのサイト](#)で確認してください

実験課題（プログラミング）

➤ P.80から順に

□ 6.11.1～6.11.6まで実施して、レポートで報告

✓ 指導書をよく読んで、必要な実験結果について記述すること

チェックポイント

➤ 6.11.1【スクリプトプログラミングの学習】

□ Pythonスクリプトの実行方法が分かっているか？

➤ 6.11.4【基本構文とファイル操作の習得】の2.

□ 最低気温、最高気温、平均気温が正しいか？

➤ 6.11.5【基本構文とコマンドライン引数の習得】の2.

□ 行などを追加しても動くかどうか？

➤ 6.11.6【総合課題】の2.

□ Excelで指定のグラフが出力できているか？

レポートについて(50点分)

➤ レポート内容(Teams課題提出, PDF形式)

□ 締め切り: 2/13(月)9:00

✓ 実験日: 12/23, 1/6, 13, 27

◆ ※提出〆切までで10点

□ ファイル名

✓ 出席番号2桁_名前_exp6.pdf

◆ 例: 00_岩田大志_exp6.pdf

- 表紙
- アブストラクト
- 実験目的(原理は不要)
- 実験結果
 - プログラムのアイデアや考えを説明する
 - 作成したプログラム(画像は禁止! コピペできるようにすること)
 - 結果(仕様と実結果を比較すること うまく動いただけではダメ)
- 参考文献 必ず記述すること
- 感想

レポート書くときの注意点

- ソースコードは適切に貼り付けてください！
 - Wordだと行間が大きくなりがちなので、12ptに固定するなど、適切に設定してください
 - ✓ オススメフォントは Courier New です
 - ✓ ただ、TeX (jlistings.sty) の方が便利ですよ！！
 - Pythonのインデントは超重要なので分かるように！！
 - ✓ できれば行番号もつけてください
 - 間違ってもソースコードを画像として貼りつけないように
 - ✓ コピペできないと検証実験できないです