First, I want to say that the following are common items the will be addressed through all chapters that are different from 1ed:

1. All examples will be updated to pandas 0.19.x
2. Examples will use python 3.x instead of 2.x
3. Examples will use Jupyter notebook instead of iPython notebook
4. Examples will all be provided locally in files, and not use wakari.io
5. There will be a slant in examples and new text towards more or a finance domain, but still focus mostly on pandas and not domain until later in the text.
6. The book will start with an intro chapter relating pandas to data science and outlining how pandas supports data science. This then allows me to rewrite some of the “word” in existing chapters to focus more on application.
7. In general, all examples will change. They may be similar, but be python 2.3, pandas 0.19.x compatible (this is actually non-trivial)
8. All chapters from 3 on will have additional / updated narrative around the data science concepts introduced in chapter 1. This may vary in extent per chapter, with likely new introductions and slightly changed wording to existing sections.

**Preface (7 pages)**

The preface gives a quick overview of the book and what we will cover. Being a 2ed, we’ll talk a little about the differences between 1ed and 2ed.

* **What is the book about?**
* **What do you need to use pandas?**
* **What assumptions are made in the book?**
* **About the second edition – why it’s different**

**Chapter 1: pandas and Data Science and Analysis (40 pages)**

The idea of this chapter is to provide some context for using pandas in the context of statistics and data science. This is often something that I think is missed in almost all pandas literature. The chapter will get into several concepts in data science and how they are supported by pandas. This will set a context for each of the subsequent chapters, mentioning each chapter with where it relates to both a) data science, and b) data science process.

* **What is pandas?**
  + **Why was it created?**
  + **What does it do?**
  + **How does it related to data science?**
* **Pandas and the data sciences**
  + **What is data science?**
  + **How does pandas relate to data science?**
  + **The pandas data science cheat sheet**
  + **The process of data science and how it is supported by pandas**
    - **retrieve data,**
    - **data preparation,**
    - **data exploration,**
    - **data modeling,**
    - **presentation**
    - **automation**
  + **Concepts of statistical analysis and application to pandas**
    - **Single and multivariate data**
    - **descriptive data,**
    - **probability,**
    - **correlation,**
    - **causation,**
    - **prediction**
  + **Types of data and applicability to pandas** 
    - **Structured**
    - **unstructured**
  + **Forms of structured data organization**
    - **categorical,**
    - **continuous,**
    - **time-series**
* **The strategy of the book**
  + **Learning pandas**
  + **Understanding how the concepts of pandas are used in data science**
  + **Closing with demonstration of data science concepts in pandas as applied to finance**

At the end of the chapter the reader will understand how each of the subsequent chapters utilized pandas to facilitate manipulating data and supporting the processes in data science.

**Chapter 2: Up and running with pandas (15 pages)**

This chapter will instruct the reader on obtain and install pandas, and to get introduce a few of the basic concepts in pandas. We will also look at how the examples are presented using iPython and Juypter notebook.

* **Installing and using pandas**
* **Using Jupyter notebooks and the source code of the book**

By the end of this chapter the user will have pandas installed and be able to use the examples in the subsequent chapters.

**Chapter 3: Representing a variable and its changes in value with the pandas Series (50 pages)**

This chapter will walk the reader through use of the pandas Series, which provides 1-diemensional, indexed, data representation. The reader will learn about how to create Series objects and to manipulate data held within. They will also learn about indexes and alignment of data, and about how the Series can be used to slice data.

* **The purpose of the Series**
* **How does a series differ from a NumPy array?**
* **Slicing a Series**
* **Creating and initializing a Series and its index**
* **Determining the shape of a Series object**
* **Heads, tails, uniqueness, and counts of values**
* **Looking up values in a Series object**
* **Boolean selection**
* **Alignment via index labels**
* **Arithmetic operations on a Series object**
* **Reindexing a Series object**
* **Applying arithmetic operations on Series objects**
* **The special case of Not-A-Number (NaN)**

By the end of this chapter the reader will have a solid understanding of using the pandas Series object.

**Chapter 4: Multi-variate data with the pandas DataFrame (70 pages)**

This chapter will walk the reader through use of the pandas DataFrame, which provides 2-diemensional, indexed, data representation. The reader will learn about how to create DataFrame objects and to manipulate data held within. They will also learn about indexes, hierarchical indexes, and alignment of data, and about how the Series can be used to slice data.

* **The purpose of a DataFrame**
* **Creating a DataFrame from scratch**
* **Loading sample data to demonstrate the capabilities of a DataFrame object**
* **Selecting columns of a DataFrame object**
* **Selecting rows and values of a DataFrame using the index**
* **Selecting rows of a DataFrame using Boolean selection**
* **Adding, replacing, and deleting columns from a DataFrame**
* **Adding, replacing, and deleting rows from a DataFrame**
* **Modifying scalar values in a DataFrame**
* **Arithmetic operations on the DataFrame objects**
* **Resetting and reindexing a DataFrame**
* **Hierarchically indexing a DataFrame**
* **Statistical methods of a DataFrame**
* **Summarized data and statistical methods of a DataFrame**

By the end of this chapter the reader will have a solid understanding of using the pandas DataFrame object.

**Chapter 5: Accessing external data sources (42 pages)**

This chapter will instruct the read in how to load and save file from different data sources into pandas data structures.

* **Why we need data from outside of pandas**
* **Reading and writing pandas data from files**
* **Working with data in CSV, JSON, HTML, Excel, and HDF5 formats**
* **Accessing data on the web and in the cloud**
* **Reading and writing from/to SQL databases**
* **Reading data from remote web data services**
* **Reading data from Google Big Table**

By the end of this chapter will have learned many of the common techniques for import and export of data from pandas.

**Chapter 6: Tidying Up Your Data (36 pages)**

This chapter will instruct the user in being able to organize data in a “tidy” form, which is usable for data analysis.

* **The concept of tidy data**
* **How pandas represents unknown values**
* **How to find NaN values in data**
* **How to filter (drop) data**
* **What pandas does with unknown values in calculations**
* **How to find, filter and fix unknown values**
* **How to identify and remove duplicate data**
* **How to transform values using replace, map, and apply**

By the end of this chapter the reader will understand the various facilities in pandas for cleaning up dirty data to make it useful for effective analysis.

**Chapter 7: Combining and Reshaping Data (30 pages)**

What the chapter is will learn how they can take data in multiple pandas object and combine those through concepts such as joins, merges and concatenation.

* **Concatenation**
* **Merging and joining and asof**
* **Pivots**
* **Stacking/unstacking**
* **Melting**
* **The potential performance benefits of stacked data**

By the end of this chapter the reader will know how to perform database-like joins and concatenation of previously disparate data sets. Taking this knowledge to the next volume, the student will then learn how to group this data and start to derive results.

**Chapter 8: Grouping and Aggregating Data (38 pages)**

In this chapter the will learn about grouping and performing aggregate analysis of data. In pandas this is often referred to as the split-apply-combine pattern. The reader will learn about using this pattern to group data in various different configurations and also to apply aggregate functions to calculate results upon each group of data.

* **An overview of the split, apply, and combine pattern for data analysis**
* **Grouping by column values**
* **Accessing the results of a grouping**
* **Grouping using index levels**
* **Applying functions to groups to create aggregate results**
* **Transforming groups of data using filtering to selectively remove groups of data**
* **The discretization of continuous data into bins**

By the end of this the reader will have a good grasp at understanding how to group data and derive results from that information.

**Chapter 9: Time-series Data (53 pages)**

This chapter will cover representing time-series data in pandas. This chapter will cover the extensive capabilities provided by pandas for facilitating analysis of this time-series data.

* **Converting string-based dates and time into objects**
* **Standardizing date and time values to specific time zones**
* **Generating sequences of fixed-frequency dates and time intervals**
* **Efficiently reading/writing the value at a specific time in a series**
* **Converting an existing time series to another with a new frequency of sampling**
* **Computing relative dates, not only taking into account time zones, but also dealing with specific calendars based upon business days**
* **Identifying missing samples in a time series and determining appropriate substitute values**
* **Shifting dates and time forward or backward by a given amount**
* **Calculating aggregate summaries of values as time changes**

By the end of this chapter the reader will understand how pandas can be used to represent complex time-related measurements in data.

**Chapter 10: Visualization (48 pages)**

This chapter will teach the reader how to create data visualizations based upon data stored in pandas data structures. We will start with the basics learning to create a simple chart from data and also control several of the attributes of the chart (such as legends, labels, colors). We will examine the creation of several common types of plots used to represent different types of data that are use those plot types to convey meaning in the underlying data. We also will learn how to integrate pandas with D3.js, so that we can create rich web based visualizations.

* **Bar plots**
* **Histograms**
* **Box and whisker charts**
* **Area plots**
* **Scatter plots**
* **Density plots**
* **Scatter plot matrixes**
* **Heatmaps**
* **D3.JS**

By the end of this chapter the reader will have an excellent grasp on how to make data visualizations using pandas, and to customize those visualizations to convey the specific meaning in the data.

**Chapter 11: Financial Applications (30 pages)**

This chapter will teach the reader will learn how to apply pandas to basic financial problems. It will focus on data obtained from Yahoo! Finance, and will demonstrate a number of financial concepts in financial data such as calculating returns, moving averages, volatility, and several other concepts. The student will also learn to apply data visualization to these financial concepts.

* **Fetching and organizing stock data from Yahoo!**
* **Plotting time-series prices**
* **Plotting volume-series data**
* **Calculating simple daily percentage change**
* **Calculating simple daily cumulative returns**
* **Resampling data from daily to monthly returns**
* **Analyzing distribution of returns**
* **Performing a moving-average calculation**
* **Comparing average daily returns across stocks**
* **Correlating stocks based on the daily percentage change of closing price**
* **Volatility calculation**
* **Determining risk relative to expected returns**

By the end of this chapter the reader will have gained the understanding of how to apply pandas to fundamental financial problems (see my other book for more detailed examples ☺). But it will also have demonstrated the concepts laid out in chapter 1 for data science relative to pandas and finance.

**Summary**

Preface: (7 Pages)

Chapter 1: pandas and Data Science and Analysis (40 pages)

Chapter 2: Up and running with pandas (15 pages)

Chapter 3: Representing a variable and its changes in value with the pandas Series (50 pages)

Chapter 4: Multi-variate data with the pandas DataFrame (70 pages)

Chapter 5: Accessing external data sources (42 pages)

Chapter 6: Tidying Up Your Data (36 pages)

Chapter 7: Combining and Reshaping Data (30 pages)

Chapter 8: Grouping and Aggregating Data (38 pages)

Chapter 9: Time-series Data (53 pages)

Chapter 10: Visualization (48 pages)

Chapter 11: Financial Applications (30 pages)

**Total Pages: 459**