



MERRIMACK COLLEGE

CSC 6000

Week 8

Basic Statistics and their Computation

Basic Programming Concepts and Discrete Mathematics - Dr. Paulo Fernandes

Presentation Agenda

Week 8

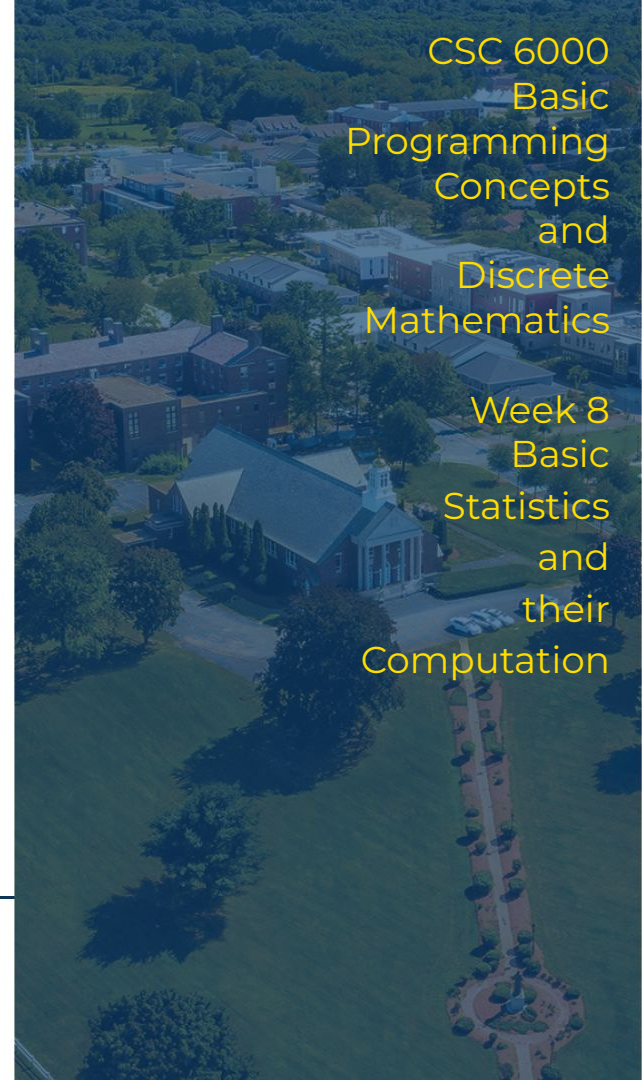
- Statistics
 - a. Statistics Basics
 - b. Sampling
 - c. Basic Measures
- Wrapping the course up
 - a. This Week's tasks
 - b. The Final Exam
 - c. Evaluation
 - d. The Next Steps



MERRIMACK COLLEGE

CSC 6000
Basic
Programming
Concepts
and
Discrete
Mathematics

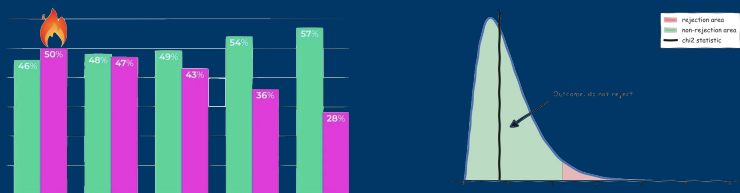
Week 8
Basic
Statistics
and
their
Computation



Statistics

Facts are stubborn things,
statistics are more pliable.

Mark Twain



There's a large number of Probability & Statistics courses available, which helps to build the confusion that the two areas are the same.

Actually, **probabilities** and **statistics** are components of the same things, just like **sowing** and **reaping**.

We will try to set these concepts apart by discussing:

- **Statistics Basics,**
- Sampling,
- Basic Measures.



Statistics

It started as an application of Probability theory, and yet the term "statistics" was written by Ghilini in 1647. Statistics became an independent area of Mathematics by the beginning of the XX Century, mostly as an effort of Galton and Pearson, establishing the basics of the area including basic measures like standard deviation. It is notable, however, that since the beginning of modern statistics, the concepts were almost always applied ones, not theoretical.

This, probably, is the reason why statistics got a bad reputation (such as Mark Twain's opinion of it).



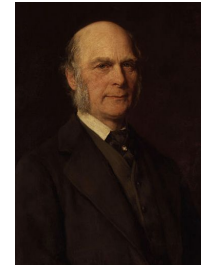
Girolamo **Ghilini**
1589 - 1632 - 1668



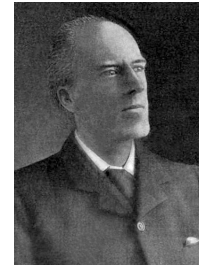
Jacob **Bernoulli**
1655 - 1684 - 1705



Carl Friedrich **Gauss**
1777 - 1840 - 1855



Francis **Galton**
1822 - 1882 - 1911



Karl **Pearson**
1857 - 1912 - 1936



MERRIMACK COLLEGE

Wikipedia: [Statistics](#).

Statistics and Probabilities

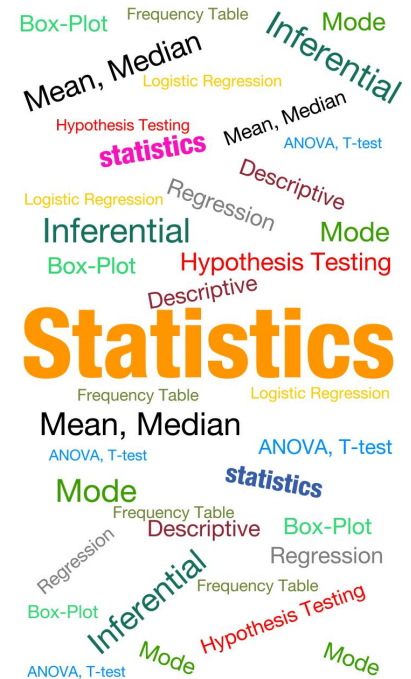
What happened?

- While:
 - in Probabilities you are trying to guess what will happen based on the likelihood of events to occur, which is fundamentally based on assumption and previous experiences,
 - in Statistics you are trying to figure out what happened in a specific phenomenon, crunching the data which you've gathered by observing some aspects of the phenomenon.
- We employ Probabilities while planning (before execution);
- We employ Statistics while analyzing (after execution).



Statistics Goals

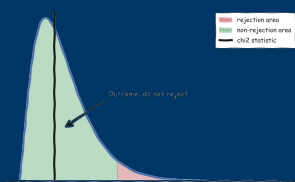
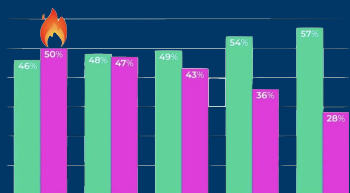
- Understanding
 - Find the proper measurements to summarize large amounts of data:
 - Data Transformations, Visualization;
 - Look for correlations among phenomena:
 - Identify Causes and Consequences, Detect Dependency;
- Credibility
 - Find numerical reasons to trust the measurements:
 - Statistical Relevance, Accountability;
 - Assign numerical values to back up opinions:
 - Assessment, Evaluation.



Statistics

Facts are stubborn things,
statistics are more pliable.

Mark Twain



The number of Probability & Statistics courses around help to build the confusion that the two areas are the same.

Actually, **probabilities** and **statistics** are components of the same things, just like **sowing** and **reaping**.

We will try to set these concepts apart by discussing:

- Statistics Basics,
- **Sampling**,
- Basic Measures.



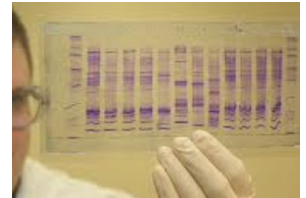
Sampling

A **population** is composed by **individuals**.
An **individual is anything**, including, but not only, people.

Collecting data

- Sometimes you have a vast amount of data.
 - Let's call it a **statistical population**, which is all data of a given phenomenon;
- If this statistical population is too big to be analyzed or even observed, we may need to set aside a portion of it to work with:
 - Let's call it a **statistical sample**, which is a (hopefully) representative subset of the statistical population.

This is traditionally what we call sampling - the techniques to pinpoint a subset of a population to be the sample to be analyzed.



Sampling

A **sample** is composed by **individuals**.
An **individual** is represented solely by its data.

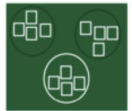
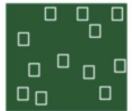
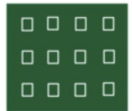
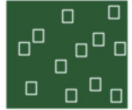
Collecting data

- Thanks to the Internet and modern technologies, the means of collecting and analyzing large amounts of data have increased significantly.
 - often all data of a given population can be collected.
- Even in such situations it is also possible to refer to it as a sample, since even when collecting all data, this can be seen as a **sample of possible data**.
 - For example, the universe that exists was just one possible instance, thus a sample of a theoretical multiverse.



Sample Characterization

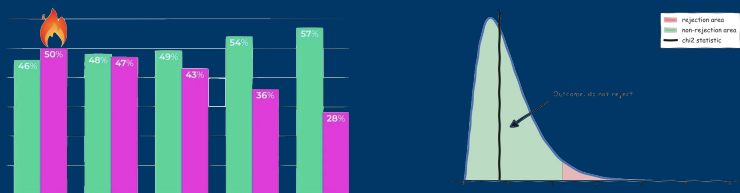
- Sampling Assumptions
 - Population Profile - the population and individuals' info
- Sampling Frame
 - Size - how many are there
 - Particular Characteristics - how are they distributed
 - Meta Information - information about the information
- Sampling Method
 - Random - just pick some
 - Convenience - just pick what is possible/easy
 - Systematic - follow a pattern
 - Cluster - pick some packs
 - Stratified - try to reproduce the population



Statistics

Facts are stubborn things,
statistics are more pliable.

Mark Twain



The number of Probability & Statistics courses around help to build the confusion that the two areas are the same.

Actually, **probabilities** and **statistics** are components of the same things, just like **sowing** and **reaping**.

We will try to set these concepts apart by discussing:

- Statistics Basics,
- Sampling,
- **Basic Measures.**

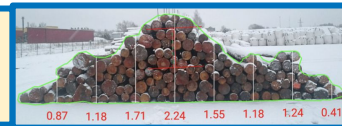


Basic Measurements

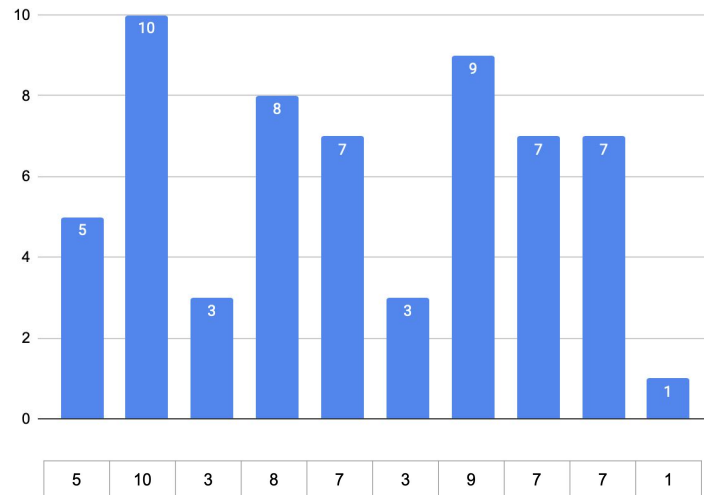
To characterize a given sample of n elements we can:

- write down all elements or ...
- compute some measurements that give the idea how the elements are:
 - the average (or mean);
 - which element is the most common one in the sample;
 - which element is the smallest one;
 - which element is the largest;
 - how spread out they are.

Given a sample composed of n values:



$n = 10$



Basic Measurements

Given a sample composed by n values.



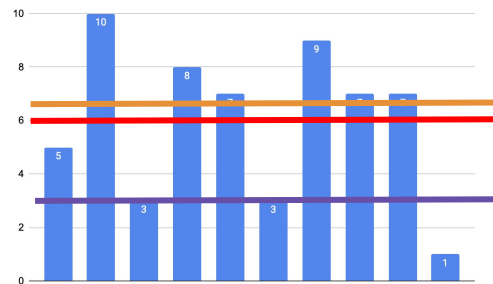
- Mean or Arithmetic Average
 - The sum of all elements of the sample divided by their number;
- Range
 - The minimum and maximum value within the sample's elements;
- Mode
 - The most frequent value within the sample's elements;
- Median
 - If sorted, the one (or average of the two) in the middle of the sample;
- Variance
 - The variability of the value with respect to the mean;
- Standard Deviation
 - The dispersion of values of the sample with respect to the mean.



MERRIMACK COLLEGE

Video: [Finding mean, median, and mode.](#)

Mean



- Mean or Arithmetic Average

- The sum of all elements of the sample divided by their number.

A =	5	10	3	8	7	3	9	7	7	1
W =	2	3	1	3	2	1	2	3	1	3

- **$n = 10$**

- average of $A = \frac{\sum_{i=0}^9 a_i}{10}$

- $\bar{A} = \frac{\sum_{i=0}^9 a_i}{10} = \frac{60}{10} = 6$

Arithmetic Average

$$\bar{A} = \frac{\sum_{i=0}^{n-1} a_i}{n}$$

- Harmonic Average **$H = 3.8003$**

$$H = \frac{n}{\sum_{i=0}^{n-1} \frac{1}{a_i}}$$

- Weighted Average

$$\bar{A}_w = \frac{\sum_{i=0}^n w_i a_i}{\sum_{i=0}^n w_i}$$

$$\bar{A}_w = 6.3333$$

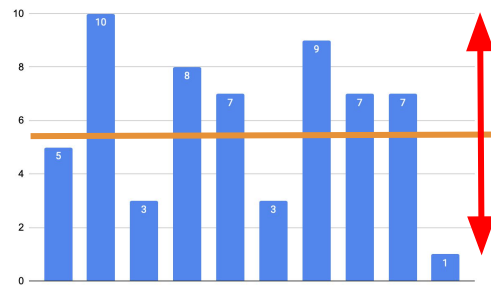


MERRIMACK COLLEGE

Wikipedia: [Average](#).

Range

- Values covered by the sample
 - The minimum and maximum value within the sample's elements.



$A =$	5	10	3	8	7	3	9	7	7	1
-------	---	----	---	---	---	---	---	---	---	---

- $n = 10$
- $\min(A) = 1$
- $\max(A) = 10$
- $\text{range}(A) = 1 - 10$

Range

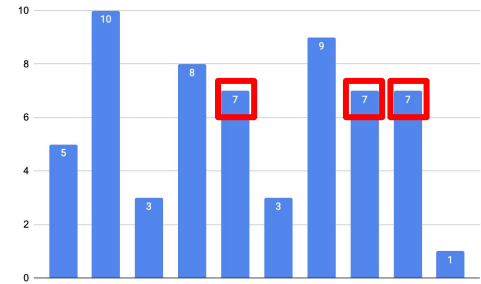
- $\min - \max$

Mid-Range

- $(\max(A) + \min(A)) / 2$
 - $(1 + 10) / 2$
 - 5.5



Mode



- The highest probability in the sample
 - The most frequent value within the sample's elements.

$A =$	5	10	3	8	7	3	9	7	7	1
-------	---	----	---	---	---	---	---	---	---	---

- **$n = 10$**
- ***the element 7 appears three times***
- *the element 3 appears two times*
- ...

Mode

- **$\text{mode}(A) = 7$**

Number of Modes

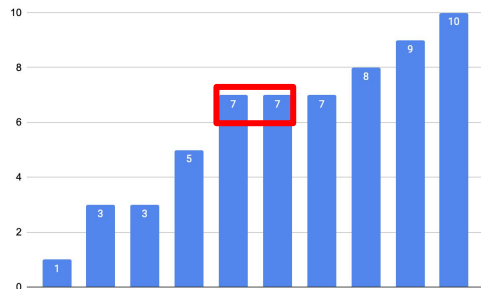
- ***Unimodal*** - 1 mode
- ***Bimodal*** - 2 modes
- ***Multimodal*** - more than 2 modes



MERRIMACK COLLEGE

Wikipedia: [Mode \(statistics\)](https://en.wikipedia.org/wiki/Mode_(statistics)).

Median



- A less stiff measure than the mean
 - If sorted, the element (or average of two) in the middle of the sample.

A =

1	3	3	5	7	7	7	8	9	10
---	---	---	---	---	---	---	---	---	----

- **$n = 10$**
- there is a even number of elements
- if sorted, the two in the middle are **7** and **7**
- **$(7 + 7) / 2 = 7$**

Median

- **$median(A) = 7$**

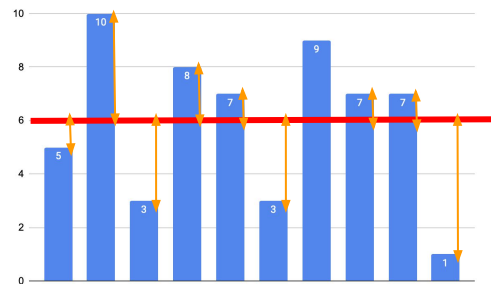
There are efficient algorithms to compute the median without sorting the sample, for example, QuickSelect



MERRIMACK COLLEGE

Wikipedia: [Median](#).

Variance (of a population)



- Basic dispersion of elements
 - The variability of the value with respect to the mean.

A =	5	10	3	8	7	3	9	7	7	1
	1	4	3	2	1	3	3	1	1	5

- **$n = 10$**
- **$\bar{A} = 6$**

- $$Var(A) = \frac{\sum_{i=1}^n (a_i - \bar{A})^2}{n}$$

Variance

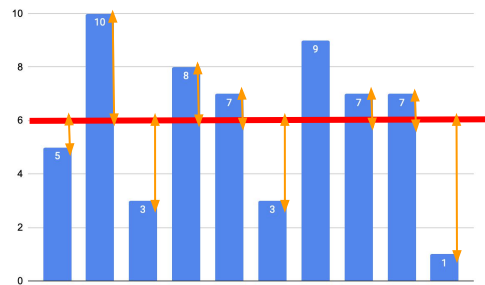
- **$var(A) = 7.6$**

The variance is the average of the square of the difference between each element and the mean.



Standard Deviation (population)

- A proportional version of the variance
 - The dispersion of values of the sample with respect to the mean.



A =

5	10	3	8	7	3	9	7	7	1
1	4	3	2	1	3	3	1	1	5

- **$n = 10$**
- **$\bar{A} = 6$**

Standard Deviation

The square root of the variance.

- $$\sigma(A) = \sqrt{\frac{\sum_{i=1}^n (a_i - \bar{A})^2}{n}}$$

- **$\sigma(A) = 2.7568$**



MERRIMACK COLLEGE

Wikipedia: [Standard Deviation](#).

Wrapping the course up

... and now the end is near.

Paul Anka (My Way lyrics)

Here we are, eight weeks after starting and (I hope) a lot more knowledgeable about Basic Programming and Discrete Mathematics.

You have learned a considerable amount of new concepts, or at least revisited concepts seen before.

Let me talk to you about what comes now and beyond:

- This Week's Tasks
- The Final Exam,
- Evaluation,
- The Next Steps.



This Week's tasks

- You have to take Quiz #8 as soon as possible (deadline is this Friday).
- Post in the discussion if the course has met your expectations.
- You have to make sure all assignments from the previous weeks are submitted, as after this Friday they will not be considered anymore.
- The final exam is available for you to take just after the last Live Session (it opens on Thursday, 8:00pm EST) and you have to take it by this Sunday at 11:59pm.



One more push and you are done!



MERRIMACK COLLEGE

Eighth Quiz - Week 8 - Q#8

- The eighth quiz in this course covers the topics of Week 8;
- The quiz is available already, and it is composed of 10 questions;
- The quiz should be taken on Canvas (Module 8), and it is not timed:
 - You can take as long as you want to answer it;
- The quiz is open book, open notes, and you can even use any language interpreter to answer it;
- Yet, the quiz is evaluated and you are allowed to submit it only once.

Your task:

- Go to Canvas, answer the quiz and submit it within the deadline:
 - The deadline for the quiz is this Friday (yes, not much time for this one).



MERRIMACK COLLEGE

This quiz counts towards
the Quizzes grade.

Post Discussion - Week 8 - D#8

Post in the discussion if the course has met your expectations:

- You can mention topics you expected to see and if they were covered in the level of detail you expected.
- You can also mention topics that you didn't expect to see, but you were positively or negatively surprised by.
- You can share your satisfaction comparing the amount of time you spent and how much you have learned.

Your task:

- Post your discussion in the message board whenever you feel like, but no later than this Friday;
- Reply to the posts of your colleagues in the message board freely.



MERRIMACK COLLEGE

This task counts towards the Discussions grade.

The Final Exam

Before starting this course, most of you took the Mastery Skill Profiler.

- You had an hour and three quarters to answer 50 questions about the very same topics covered here in class.
- The final exam is your chance to prove to yourselves how well you did over the last eight weeks.



Now as the final task of this course, you have to take the final exam. It is very similar to the Profiler, and all the quizzes.

- The main difference is that it is timed and you have just 75 minutes to answer the 50 questions.
 - The questions might be a little, very, or not at all similar to those on the Profiler and your quizzes.



MERRIMACK COLLEGE

This task counts towards the Final Exam grade.

Evaluation

The evaluated tasks are weighted as such (103%):

Activity	Published grades	Percentage
Projects (7)	Evaluated from 0 to 100 each	42.0%
Quizzes (8)	Evaluated from 0 to 10 each	32.0%
Discussions (8)	Evaluated from 0 to 100 each	4.0%
Final Exam	Evaluated from 0 to 50	25.0%

All tasks are due by their stated deadline, the late penalties reduce 10% of the evaluation, plus another 2% for each full day of delay.



MERRIMACK COLLEGE

Final deadline on Friday of the course's last week

Evaluation

The final letter grade is computed according to this table:

A	A-	B+	B	B-	C+	C	C-	F
95 and up	90 to 94.9	87 to 89.9	83 to 86.9	80 to 82.9	77 to 79.9	73 to 76.9	70 to 72.9	69.9 and low

There are no grade D+, D, or D- in graduate courses. C- is not a passing grade for MSCS. An average grade B is required to remain in the Masters of Science in Computer Science, an average lower than that puts you in probation.



MERRIMACK COLLEGE

Final grades published first Monday after the course's end.

The Next Steps

Now that you finished the course, successfully I hope, this is your path in the MSCS Program:

- CSC 6003 Foundations of Programming
- CSC 6013 Algorithms & Discrete Structures
- CSC 6023 Advanced Algorithms
- CSC 6033 Automata, Languages, and Computability
 - DSE 6003 Data Governance and Privacy
- CSC 6301 Software Design & Documentation
- CSC 6302 Database Principles
- CSC 6303 Systems & Languages Survey
- CSC 6304 Advanced Programming Concepts
- CSC 6400 Capstone Project (optional)



```
80 return array(['code' => $captcha_config['code'], 'image_src' => $image_src]);
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
```



MERRIMACK COLLEGE

Keep in touch with the success coach!

” **This was CSC 6000,
my pleasure and honor to be your instructor!**

- **Post discussion D#8, like, now;**
- **Take Quiz Q#8, no later than this Friday;**
- **Take the Final exam no later than this Sunday.**

Next Week - Foundations of Programming!



MERRIMACK COLLEGE